PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07H 21/02, 21/04, 1/00, 14/00, 17/00, C12Q 1/68, G01N 33/53

A1

(11) International Publication Number: WO 99/06426

(43) International Publication Date: 11 February 1999 (11.02.99)

(21) International Application Number: PCT/US98/16102

(22) International Filing Date: 3 August 1998 (03.08.98)

(30) Priority Data:
60/054,646 4 August 1997 (04.08.97) US
60/091,650 2 July 1998 (02.07.98) US

(71) Applicant: MILLENNIUM BIOTHERAPEUTICS, INC. [US/US]; 620 Memorial Drive, Cambridge, MA 02142 (US).

(72) Inventor: PAN, Yang; 6 Hamilton Road #1, Brookline, MA 02146 (US).

(74) Agent: MEIKLEJOHN, Anita, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NOVEL MOLECULES OF THE TANGO-77 RELATED PROTEIN FAMILY AND USES THEREOF

(57) Abstract

Novel **Tango-77** polypeptides, proteins, and nucleic acid molecules are disclosed. In addition to isolated, full-length Tango-77 proteins, the invention further provides isolated Tango-77 fusion proteins, antigenic peptides and anti-Tango-77 antibodies. The invention also provides Tango-77 nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecule of the invention, host cells into which the expression vectors have been introduced and non-human transgenic animals in which a Tango-77 gene has been introduced or disrupted. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

NOVEL MOLECULES OF THE TANGO-77 RELATED PROTEIN FAMILY AND USES THEREOF

Background of the Invention

The polypeptide cytokine interleukin-1 (IL-1) is a critical mediator of inflammatory and overall immune response. To date, three members of the IL-1 family, IL-1α, IL-1β and IL-1ra (Interleukin-1 receptor antagonist) have been isolated and cloned. IL-1α and IL-1β are proinflammatory cytokines which elicit biological responses, whereas IL-1ra is an antagonist of IL-1α and IL-1β activity. Two distinct cell-surface receptors have been identified for these ligands, the type 1 IL-1 receptor (IL-1RtI) and type II IL-1 receptor (IL-1RtII). Recent results suggest that the IL-1RtI is the receptor responsible for transducing a signal and producing biological effects.

As mentioned above, IL-1 is a key mediator of the host inflammatory response. While inflammation is an important homeostatic mechanism, aberrant inflammation has the potential for inducing damage to the host. Elevated IL-1 levels are known to be associated with a number of diseases particularly autoimmune diseases and inflammatory disorders.

Since Il-1ra is a naturally occurring inhibitor of IL-1, IL-1ra can be used to limit the aberrant and potentially deleterious effects of IL-1. In experimental animals, pretreatment with IL-1ra has been shown to prevent death resulting from lipopolysaccharide-induced sepsis. The relative absence of IL-1ra has also been suggested to play a role in human inflammatory bowel disease.

Summary of the Invention

The present invention is based, at least in part, on the discovery of a gene encoding Tango-77, a secreted

protein that is predicted to be a member of the cytokine superfamily. The Tango-77 cDNA described below (SEQ ID NO:1) has three possible open reading frames. The first potential open reading frame encompasses 534 nucleotides extending from nucleotide 356 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:3) and encodes a 178 amino acid protein (SEQ ID NO:2). This protein may include a predicted signal sequence of about 63 amino acids (from about amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID NO:4) and a predicted mature protein of about 115 amino acids (from about amino acid 64 to amino acid 178 of SEQ ID NO:2 (SEQ ID NO:5)).

The second potential open reading frame encompasses 498 nucleotides extending from nucleotide 389 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:6) and encodes a 167 amino acid protein (SEQ ID NO:7). This protein may include a predicted signal sequence of about 52 amino acids (from about amino acid 1 to about amino acid 52 of SEQ ID NO:7 (SEQ ID NO:8)) and a predicted mature protein of about 115 amino acids (from about amino acid 52 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:9)).

The third potential open reading frame encompasses 408 nucleotides extending from nucleotide 481 to nucleotide 889 of SEQ ID NO:1 (SEQ ID NO:10) and encodes a 136 amino acid protein (SEQ ID NO:11). This protein includes a predicted signal sequence of about 21 amino acids (from about amino acid 1 to about amino acid 21 of SEQ ID NO:11 (SEQ ID NO:12)) and a predicted mature protein of about 115 amino acids (from about amino acid 22 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:13)).

As used herein, the terms "Tango-77", "Tango-77 protein", "Tango-77 polypeptide" amd the like, can refer and polypeptide produced by the cDNA of SEQ ID NO:1 including any and all of the Tango-77 gene products described above.

Tango-77 is expected to inhibit inflammation and play a functional role similar to that of secreted IL-1ra. For example, it is expected that Tango-77 may bind to the IL-1 receptor, thus blocking receptor activation by inhibiting the binding of IL-1α and IL-1β to the receptor. Alternatively, Tango-77 may inhibit inflammation through another pathway, for example, by binding to a novel receptor. Accordingly, Tango-77 may be useful as a modulating agent in regulating a variety of cellular processes including acute and chronic inflammation, e.g., asthma, chronic myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory bowel disease.

In one aspect, the invention provides isolated nucleic acid molecules encoding Tango-77 or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of Tango-77.

The invention encompasses methods of diagnosing and treating patients who are suffering from a disorder associated with an abnormal level (undesirably high or undesirably low) of inflammation, abnormal activity of the IL-1 receptor complex, or abnormal activity of IL-1, by administering a compound that modulates the expression of Tango-77 (at the DNA, mRNA or protein level, e.g., by altering mRNA splicing) or by altering the activity of Tango-77. Examples of such compounds include small molecules, antisense nucleic acid molecules, ribozymes, and polypeptides.

The invention features a nucleic acid molecule which is at least 45% (e.g., 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the nucleotide sequence of the cDNA insert of the plasmid

deposited with ATCC as Accession Number (the "cDNA of ATCC 98807"), or a complement thereof.

The invention features a nucleic acid molecule which includes a fragment of at least 100 (e.g., 250, 325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, or 989) nucleotides of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the nucleotide sequence of the cDNA ATCC 98807, or a complement thereof.

The invention also features a nucleic acid molecule which includes a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, or the amino acid sequence encoded by the cDNA of ATCC 98807.

In a preferred embodiment, a Tango-77 nucleic acid molecule has the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the nucleotide sequence of the cDNA of ATCC 98807.

Also within the invention is a nucleic acid molecule which encodes a fragment of a polypeptide having the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, wherein the fragment includes at least 15 (e.g., 25, 30, 50, 100, 150, or 178) contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or the polypeptide encoded by the cDNA of ATCC Accession Number 98807.

The invention includes a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or

an amino acid sequence encoded by the cDNA of ATCC Accession Number 98807, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or a complement thereof under stringent conditions.

Also within the invention are: an isolated Tango-77 protein having an amino acid sequence that is at least about 45%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:5, SEQ ID NO:9 or SEQ ID NO:13 (mature human Tango-77), or the amino acid sequence of SEQ ID NO:2, SEQ ID NO:7 or SEQ ID NO:11 (immature human Tango-77).

Also within the invention are: an isolated Tango-77 protein which is encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 65%, preferably 75%, 85%, or 95% identical to SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the cDNA of ATCC 98807; and an isolated Tango-77 protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the non-coding strand of the cDNA of ATCC 98807, or the complement thereof.

Also within the invention is a polypeptide which is a naturally occurring allelic variant of a polypeptide that includes the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID

NO:10 or the complement thereof under stringent conditions.

Another embodiment of the invention features Tango-77 nucleic acid molecules which specifically detect Tango-77 nucleic acid molecules relative to nucleic acid molecules encoding other members of the cytokine superfamily. For example, in one embodiment, a Tango-77 nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement thereof. In another embodiment, the Tango-77 nucleic acid molecule is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, or 989) nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement thereof. In yet another embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a Tango-77 nucleic acid.

Another aspect of the invention provides a vector, e.g., a recombinant expression vector, comprising a Tango-77 nucleic acid molecule of the invention. In another embodiment, the invention provides a host cell containing such a vector. The invention also provides a method for producing Tango-77 protein by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a Tango-77 protein is produced.

Another aspect of this invention features isolated or recombinant Tango-77 proteins and polypeptides. Preferred Tango-77 proteins and polypeptides possess at least one biological activity possessed by naturally

occurring human Tango-77, e.g., (i) the ability to interact with proteins in the Tango-77 signalling pathway (ii) the ability to interact with a Tango-77 ligand or receptor; or (iii) the ability to interact with an intracellular target protein, (iv) the ability to interact with a protein involved in inflammation and (v) the ability to bind the IL-1 receptor. Other activities include the induction and suppression of polypeptide interleukins, cytokines and growth factors.

The Tango-77 proteins of the present invention, or biologically active portions thereof, can be operably linked to a non-Tango-77 polypeptide (e.g., heterologous amino acid sequences) to form Tango-77 fusion proteins. The invention further features antibodies that specifically bind Tango-77 proteins, such as monoclonal or polyclonal antibodies. In addition, the Tango-77 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of Tango-77 activity or expression in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of Tango-77 activity or expression such that the presence of Tango-77 activity or expression is detected in the biological sample.

In another aspect, the invention provides a method for modulating Tango-77 activity comprising contacting a cell with an agent that modulates (inhibits or stimulates) Tango-77 activity or expression such that Tango-77 activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to Tango-77 protein. In another embodiment, the

agent modulates expression of Tango-77 by modulating transcription of a Tango-77 gene, splicing of a Tango-77 mRNA, or translation of a Tango-77 mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the Tango-77 mRNA or the Tango-77 gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant Tango-77 protein activity or nucleic acid expression by administering an agent which is a Tango-77 modulator to the subject. In one embodiment, the Tango-77 modulator is a Tango-77 protein. In another embodiment, the Tango-77 modulator is a Tango-77 nucleic acid molecule. In other embodiments, the Tango-77 modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant Tango-77 protein or nucleic acid expression can include chronic and acute inflammation.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a Tango-77 protein; (ii) mis-regulation of a gene encoding a Tango-77 protein; and (iii) aberrant post-translational modification of a Tango-77 protein, wherein a wild-type form of the gene encodes a protein with a Tango-77 activity.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a Tango-77 protein. In general, such methods entail measuring a biological activity of a Tango-77 protein in the presence and absence of a test compound and identifying those

compounds which alter the activity of the Tango-77 protein.

The invention also features methods for identifying a compound which modulates the expression of Tango-77 by measuring the expression of Tango-77 in the presence and absence of a compound.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence (SEQ ID NO:1) of Tango-77. The Tango-77 cDNA has three possible open reading frames which encode the amino acid sequence (SEQ ID NO:2, SEQ ID NO:7 and SEQ ID NO:11) of human Tango-77. The three potential open reading frames of SEQ ID NO:1 extend from: (1) nucleotide 356 to nucleotide 889 (SEQ ID NO:3); (2) nucleotide 389 to nucleotide 889 (SEQ ID NO:6); and (3) nucleotide 481 to nucleotide 889 (SEQ ID NO:10).

Figure 2 depicts an alignment of an amino acid sequence of Tango-77 (T77; SEQ ID NO:2) with IL-1RA (SEQ ID NO:14), and IL-1β (SEQ ID NO:15).

Figure 3 depicts the genomic sequence of BAC1 (SEQ ID NO:16).

Figure 4 depicts the genomic sequence of BAC2 (SEQ ID NO:17).

Figure 5 depicts an amino acid sequence of an alternatively spliced form of Tango-77 (SEQ ID NO:2) as predicted by Procrustes (T77-procrustes; SEQ ID NO:18).

Figure 6 depicts an alignment of an amino acid sequence of an alternatively spliced form of Tango-77 (T77-procrustes; SEQ ID NO:18) with Tango-77 (SEQ ID NO:2).

Figure 7 depicts an alignment of an amino acid sequence of an alternatively spliced form of Tango-77 (T77-procrustes; SEQ ID NO:18) with IL-1ra (SEQ ID NO:14), and IL-1β (SEQ ID NO:15).

Detailed Description of the Invention

The present invention is based on the discovery of a cDNA molecule encoding human Tango-77, a member of the cytokine superfamily. The cDNA molecule encoding human Tango-77 has three possible open reading frames. The three possible nucleotide open reading frames for human Tango-77 protein are shown in Figure 1 (SEQ ID NO:3, SEQ ID NO:6 and SEQ ID NO:10). The predicted amino acid sequence for the three possible Tango-77 immature proteins are also shown in Figure 1 (SEQ ID NO:2, SEQ ID NO:7 or SEQ ID NO:11) and three possible mature proteins are also shown in Figure 1 (SEQ ID NO:5, SEQ ID NO:9 and SEQ ID NO:13).

The Tango-77 cDNA of Figure 1 (SEQ ID NO:1), which is approximately 989 nucleotides long including untranslated regions, encodes a protein amino acid having a molecular weight of approximately 19 kDa, 18 kDa, or 14.9 KDa (excluding post-translational modifications) and the possible mature form of the protein has a molecular weight of 13 kDa. A plasmid containing a cDNA encoding human Tango-77 (with the cDNA insert name of Of fthx077) was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia 20110-2209 on July 2, 1998 and assigned Accession Number 98807. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an

admission that a deposit is required under 35 U.S.C. §112.

Human Tango-77 is one member of a family of molecules (the "Tango-77 family") having certain conserved structural and functional features. The term "family," when referring to the protein and nucleic acid molecules of the invention, is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin and a homologue of that protein of murine origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics.

As used interchangeably herein a "Tango-77 activity", "biological activity of Tango-77" or "functional activity of Tango-77", refers to an activity exerted by a Tango-77 protein, polypeptide or nucleic acid molecule on a Tango-77 responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. A Tango-77 activity can be a direct activity, such as an association with a second protein, or an indirect activity, such as a cellular signaling activity mediated by interaction of the Tango-77 protein with a second protein. In a preferred embodiment, a Tango-77 activity includes at least one or more of the following activities: (i) the ability to interact with proteins in the Tango-77 signalling pathway (ii) the ability to interact with a Tango-77 ligand or receptor; or (iii) the ability to interact with an intracellular target protein, (iv) the ability to interact with a protein involved in

inflammation, and (v) the ability to bind the IL-1 receptor.

Accordingly, another embodiment of the invention features isolated Tango-77 proteins and polypeptides having a Tango-77 activity.

Yet another embodiment of the invention features Tango-77 molecules which contain a signal sequence. Generally, a signal sequence (or signal peptide) is a peptide containing about 21 to 63 amino acids which occurs at the extreme N-terminal end of a secretory protein. The native Tango-77 signal sequence (SEQ ID NO:4, SEQ ID NO:8, or SEQ ID NO:12) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of Tango-77 can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence. Alternatively, the native Tango-77 signal sequence can itself be used as a heterologous signal sequence in expression systems, e.g., to facilitate the secretion of a protein of interest.

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode Tango-77 proteins or biologically active portions thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify Tango-77-encoding nucleic acids (e.g., Tango-77 mRNA) and fragments for use as PCR primers for the amplification or mutation of Tango-77 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g.,

cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5′ and 3′ ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated Tango-77 nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or the complement thereof as a hybridization probe, Tango-77 nucleic acid molecules can be isolated using standard

hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to Tango-77 nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 the cDNA of ATCC 98807, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding Tango-77, for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of Tango-77. The nucleotide sequence determined from the cloning of the human Tango-77 gene allows for the generation of probes and primers designed for use in identifying and/or cloning Tango-77 homologues in other cell types, e.g., from other tissues, as well as Tango-77 homologues from other mammals. The probe/primer typically comprises

substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807. Alternatively, the oligonucleotide can typically comprise a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of a naturally occurring mutant of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

Probes based on the human Tango-77 nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or identical proteins. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a Tango-77 protein, such as by measuring a level of a Tango-77-encoding nucleic acid in a sample of cells from a subject, e.g., detecting Tango-77 mRNA levels or determining whether a genomic Tango-77 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of Tango-77" can be prepared by isolating a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the nucleotide sequence of the cDNA of ATCC 98807 which encodes a polypeptide having a Tango-77 biological activity, expressing the encoded portion of Tango-77 protein (e.g., by recombinant expression *in*

vitro) and assessing the activity of the encoded portion of Tango-77.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807 due to degeneracy of the genetic code and thus encode the same Tango-77 protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

In addition to the human Tango-77 nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of Tango-77 may exist within a population (e.g., the human population). Such genetic polymorphism in the Tango-77 gene may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. As used herein, the term "allelic variant" refers to a nucleotide sequence which occurs at a Tango-77 locus or to a polypeptide encoded by the nucleotide sequence. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a Tango-77 protein, preferably a mammalian Tango-77 protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the Tango-77 gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations in

Tango-77 that are the result of natural allelic variation and that do not alter the functional activity of Tango-77 are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding Tango-77 proteins from other species (Tango-77 homologues), which have a nucleotide sequence which differs from that of a human Tango-77, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the Tango-77 cDNA of the invention can be isolated based on their identity to the human Tango-77 nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, or 989) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule

of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA of ATCC 98807, or the complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the Tango-77 sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the cDNA of ATCC 98807, thereby leading to changes in the amino acid sequence of the encoded Tango-77 protein, without altering the biological activity of the Tango-77 protein. Amino acid residues that are not conserved or only semiconserved among Tango-77 of various species may be non-essential for activity and thus would likely be targets for alteration. Alternatively, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of Tango-77 (e.g., the sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the Tango-77 proteins of various species may be essential for activity and thus would not likely be targets for alteration, unless one wishes to reduce or alter Tango-77 activity.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding Tango-77

proteins that contain changes in amino acid residues that are not essential for activity. Such Tango-77 proteins differ in amino acid sequence from SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13.

An isolated nucleic acid molecule encoding a Tango-77 protein having a sequence which differs from that of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the cDNA of ATCC 98807 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine,

valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in Tango-77 is preferably replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of a Tango-77 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for Tango-77 biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant Tango-77 protein can be assayed for: (1) the ability to form protein:protein interactions with proteins in the Tango-77 signalling pathway; (2) the ability to bind a Tango-77 ligand or receptor; or (3) the ability to bind to an intracellular target protein or (4) the ability to interact with a protein involved in inflammation or (5) the ability to bind the IL-1 receptor. In yet another preferred embodiment, a mutant Tango-77 can be assayed for the ability to modulate inflammation, asthma, autoimmune dieseases, and sepsis.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire Tango-77 coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading

frame). An antisense nucleic acid molecule can be antisense to a noncoding region of the coding strand of a nucleotide sequence encoding Tango-77. The noncoding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

Given the coding strand sequences encoding Tango-77 disclosed herein (e.g.,SEQ ID NO:3, SEQ ID NO:5, or SEQ ID NO:8), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of Tango-77 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of Tango-77 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of Tango-77 mRNA, e.g., an oligonucleotide having the sequence 5'-TGCAACTTTTACAGGAAACAC-3' (SEQ ID NO:19) or 5'-CCTCACTTTTACCCGAGACTC-3' (SEQ ID NO:20) or 5'-GACGGGTGGTACTTAAAACAA-3' (SEQ ID NO:21). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to

generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a Tango-77 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which

binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead

ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave Tango-77 mRNA transcripts to thereby inhibit translation of Tango-77 mRNA. A ribozyme having specificity for a Tango-77-encoding nucleic acid can be designed based upon the nucleotide sequence of a Tango-77 cDNA disclosed herein (e.g., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Tango-77-encoding mRNA. *See, e.g.*, Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, Tango-77 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, Tango-77 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the Tango-77 (e.g., the Tango-77 promoter and/or enhancers) to form triple helical structures that prevent transcription of the Tango-77 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des*. 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci*. 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see Hyrup et al*. (1996) *Bioorganic*

& Medicinal Chemistry 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs of Tango-77 can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs of Tango-77 can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996) *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996) *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs of Tango-77 can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of Tango-77 can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion

would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) *supra* and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acid Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.*, Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide,

hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Tango-77 Proteins and Anti-Tango-77 Antibodies

One aspect of the invention pertains to isolated Tango-77 proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-Tango-77 antibodies. In one embodiment, native Tango-77 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, Tango-77 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a Tango-77 protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the Tango-77 protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of Tango-77 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, Tango-77 protein that is substantially free of cellular material includes preparations of Tango-77 protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of non-Tango-77 protein (also referred to herein as a "contaminating protein"). When the Tango-77 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture

medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When Tango-77 protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of Tango-77 protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or non-Tango-77 chemicals.

Biologically active portions of a Tango-77 protein include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the Tango-77 protein (e.g., the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13), which include fewer amino acids than the full length Tango-77 proteins, and exhibit at least one activity of a Tango-77 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the Tango-77 protein. A biologically active portion of a Tango-77 protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native Tango-77 protein.

Preferred Tango-77 protein has the amino acid sequence shown of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13. Other useful Tango-77 proteins are substantially identical to SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 and retain the functional activity of the protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 yet differ in

amino acid sequence due to natural allelic variation or mutagenesis. Accordingly, a useful Tango-77 protein is a protein which includes an amino acid sequence at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99% identical to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13 and retains the functional activity of the Tango-77 proteins of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13. In a preferred embodiment, the Tango-77 protein retains a functional activity of the Tango-77 protein of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, or SEQ ID NO:13.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions, e.g., overlapping x 100). Preferably, the two sequences are the same length.

The determination of percent homology between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990)

Proc. Natl. Acad. Sci. USA 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to Tango-77 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to Tango-77 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides Tango-77 chimeric or fusion proteins. As used herein, a Tango-77 "chimeric protein" or "fusion protein" comprises a Tango-77 polypeptide operably linked to a non-Tango-77 polypeptide. A "Tango-77 polypeptide" refers to a

polypeptide having an amino acid sequence corresponding to Tango-77 polypeptides, whereas a "non-Tango-77 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially identical to the Tango-77 protein, e.g., a protein which is different from the Tango-77 protein and which is derived from the same or a different organism. Within a Tango-77 fusion protein the Tango-77 polypeptide can correspond to all or a portion of a Tango-77 protein, preferably at least one biologically active portion of a Tango-77 protein. Within the fusion protein, the term "operably linked" is intended to indicate that the Tango-77 polypeptide and the non-Tango-77 polypeptide are fused in-frame to each other. The non-Tango-77 polypeptide can be fused to the N-terminus or C-terminus of the Tango-77 polypeptide.

One useful fusion protein is a GST-Tango-77 fusion protein in which the Tango-77 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant Tango-77.

In another embodiment, the fusion protein is a Tango-77 protein containing a heterologous signal sequence at its N-terminus. For example, the native Tango-77 signal sequence (i.e., about amino acids 1 to 63 of SEQ ID NO:2; SEQ ID NO:4; or about amino acids 1 to 52 of SEQ ID NO:7;SEQ ID NO:8; or about amino acids 1 to 21 of SEQ ID NO:11;SEQ ID NO:12) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of Tango-77 can be increased through use of a heterologous signal sequence. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Ausubel et al., supra*). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of

melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (*Sambrook et al.*, *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an Tango-77-immunoglobulin fusion protein in which all or part of Tango-77 is fused to sequences derived from a member of the immunoglobulin protein family. The Tango-77-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a Tango-77 ligand and a Tango-77 receptor on the surface of a cell, to thereby suppress Tango-77-mediated signal transduction *in vivo*. The Tango-77-immunoglobulin fusion proteins can be used to affect the bioavailability of a Tango-77 cognate ligand. Inhibition of the Tango-77 ligand/Tango-77 interaction may be useful therapeutically for both the treatment of inflammatory and autoimmune disorders. Moreover, the Tango-77-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-Tango-77 antibodies in a subject, to purify Tango-77 ligands and in screening assays to identify molecules which inhibit the interaction of Tango-77 with a Tango-77 receptor.

Preferably, a Tango-77 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid

undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g., Current Protocols in Molecular Biology*, Ausubel et al. eds., John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). An Tango-77-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the Tango-77 protein.

The present invention also pertains to variants of the Tango-77 proteins (i.e., proteins having a sequence which differs from that of the Tango-77 amino acid sequence). Such variants can function as either Tango-77 agonists (mimetics) or as Tango-77 antagonists. Variants of the Tango-77 protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the Tango-77 protein. An agonist of the Tango-77 protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the Tango-77 protein. An antagonist of the Tango-77 protein can inhibit one or more of the activities of the naturally occurring form of the Tango-77 protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the Tango-77 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer

side effects in a subject relative to treatment with the naturally occurring form of the Tango-77 proteins.

Variants of the Tango-77 protein which function as either Tango-77 agonists (mimetics) or as Tango-77 antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the Tango-77 protein for Tango-77 protein agonist or antagonist activity. In one embodiment, a variegated library of Tango-77 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of Tango-77 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential Tango-77 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of Tango-77 sequences therein. There are a variety of methods which can be used to produce libraries of potential Tango-77 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential Tango-77 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the Tango-77 protein coding sequence can be used to generate a variegated population of Tango-77 fragments for

screening and subsequent selection of variants of a Tango-77 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a Tango-77 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the Tango-77 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of Tango-77 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify Tango-77 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated Tango-77 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind Tango-77 using standard techniques for polyclonal and monoclonal antibody preparation. The full-length Tango-77 protein can be used or, alternatively, the invention provides antigenic peptide fragments of Tango-77 for use as immunogens. The antigenic peptide of Tango-77 comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11 or SEQ ID NO:13 and encompasses an epitope of Tango-77 such that an antibody raised against the peptide forms a specific immune complex with Tango-77.

A Tango-77 immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed Tango-77 protein or a chemically synthesized Tango-77 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic Tango-77 preparation induces a polyclonal anti-Tango-77 antibody response.

Accordingly, another aspect of the invention pertains to anti-Tango-77 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as Tango-77. A molecule which specifically binds to Tango-77 is a molecule which binds Tango-77, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains Tango-77.

Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind Tango-77. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of Tango-77. A monoclonal antibody composition thus typically displays a single binding affinity for a particular Tango-77 protein with which it immunoreacts.

Polyclonal anti-Tango-77 antibodies can be prepared as described above by immunizing a suitable subject with a Tango-77 immunogen. The anti-Tango-77 antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized Tango-77. If desired, the antibody molecules directed against Tango-77 can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-Tango-77 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally* Current

Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a Tango-77 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds Tango-77.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-Tango-77 monoclonal antibody (*see, e.g.*, Current Protocols in Immunology, *supra*; Galfre et al. (1977) *Nature* 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) *Yale J. Biol. Med.*, 54:387-402. Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line, e.g., a myeloma cell line that is sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused

myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind Tango-77, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-Tango-77 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with Tango-77 to thereby isolate immunoglobulin library members that bind Tango-77. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734.

Additionally, recombinant anti-Tango-77 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art,

for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heave and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of Tango-77. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and

human monoclonal antibodies and protocols for producing such antibodies, see, e.g., U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to the described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope.

First, a non-human monoclonal antibody which binds a selected antigen (epitope), e.g., an antibody which inhibits Tango-77 activity, is identified. The heave chain and the light chain of the non-human antibody are cloned and used to create phage display Fab fragments. For example, the heave chain gene can be cloned into a plasmid vector so that the heavy chain can be secreted from bacteria. The light chain gene can be cloned into a phage coat protein gene so that the light chain can be expressed on the surface of phage. A repertoire (random collection) of human light chains fused to phage is used to infect the bacteria which express the non-human heavy chain. The resulting progeny phage display hybrid antibodies (human light chain/non-human heavy chain). The selected antigen is used in a panning screen to select phage which bind the selected antigen. Several rounds of selection may be required to identify such phage. Next, human light chain genes are isolated from the selected phage which bind the selected antigen. These selected human light chain genes are then used to guide the selection of human heavy chain genes as

follows. The selected human light chain genes are inserted into vectors for expression by bacteria. Bacteria expressing the selected human light chains are infected with a repertoire of human heavy chains fused to phage. The resulting progeny phage display human antibodies (human light chain/human heavy chain).

Next, the selected antigen is used in a panning screen to select phage which bind the selected antigen. The phage selected in this step display completely human antibody which recognize the same epitope recognized by the original selected, non-human monoclonal antibody. The genes encoding both the heavy and light chains are readily isolated and be further manipulated for production of human antibody. This technology is described by Jespers et al. (1994, *Bio/technology* 12:899-903).

An anti-Tango-77 antibody (e.g., monoclonal antibody) can be used to isolate Tango-77 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-Tango-77 antibody can facilitate the purification of natural Tango-77 from cells and of recombinantly produced Tango-77 expressed in host cells. Moreover, an anti-Tango-77 antibody can be used to detect Tango-77 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the Tango-77 protein. Anti-Tango-77 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials.

Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid molecule encoding Tango-77 (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant

DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression

vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., Tango-77 proteins, mutant forms of Tango-77, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of Tango-77 in prokaryotic or eukaryotic cells, e.g., bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia

Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al. (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the Tango-77 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1

(Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, Tango-77 can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al. (*supra*).

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore

(1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to Tango-77 mRNA. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see

Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, Tango-77 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome.

In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding Tango-77 or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) Tango-77 protein. Accordingly, the invention further provides methods for producing Tango-77 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding Tango-77 has been introduced) in a suitable medium such that Tango-77 protein is produced. In another embodiment, the method further comprises isolating Tango-77 from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which Tango-77-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous Tango-77 sequences have been introduced into their genome or homologous recombinant animals in which endogenous Tango-77 sequences have been altered. Such animals are useful for

studying the function and/or activity of Tango-77 and for identifying and/or evaluating modulators of Tango-77 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous Tango-77 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing Tango-77-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The Tango-77 cDNA sequence e.g., that of (SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6; SEQ ID NO:10 or the cDNA of ATCC 98807) can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human Tango-77 gene, such as a mouse Tango-77 gene, can be isolated based on hybridization to the human Tango-77 cDNA and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency

of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the Tango-77 transgene to direct expression of Tango-77 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the Tango-77 transgene in its genome and/or expression of Tango-77 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding Tango-77 can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a Tango-77 gene (e.g., a human or a non-human homolog of the Tango-77 gene, e.g., a murine Tango-77 gene) into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the Tango-77 gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous Tango-77 gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous Tango-77 gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby

alter the expression of the endogenous Tango-77 protein). In the homologous recombination vector, the altered portion of the Tango-77 gene is flanked at its 5′ and 3′ ends by additional nucleic acid of the Tango-77 gene to allow for homologous recombination to occur between the exogenous Tango-77 gene carried by the vector and an endogenous Tango-77 gene in an embryonic stem cell. The additional flanking Tango-77 nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5′ and 3′ ends) are included in the vector (*see, e.g.*, Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced Tango-77 gene has homologously recombined with the endogenous Tango-77 gene are selected (*see, e.g.*, Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (*see, e.g.*, Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.*, Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The Tango-77 nucleic acid molecules, Tango-77 proteins, and anti-Tango-77 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, (e.g. intravenous, intradermal, subcutaneous) (e.g., oral inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the

adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including

in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a Tango-77 protein or anti-Tango-77 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a

glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be

obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (*see, e.g.*, Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can

include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). A Tango-77 protein interacts with other cellular proteins and can thus be used for regulation of inflammation. The polypeptides of the invention can be used in assays to determine biological activity. For example, they could be used in a panel of proteins for high-throughput screening.

The isolated nucleic acid molecules of the invention can be used to express Tango-77 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect Tango-77 mRNA (e.g., in a biological sample) or a genetic lesion in a Tango-77 gene, and to modulate Tango-77 activity. In addition, the Tango-77 proteins can be used to screen drugs or compounds which modulate the Tango-77 activity or expression as well as to treat disorders characterized by insufficient or excessive production of Tango-77 protein or production of Tango-77 protein forms which have decreased or aberrant activity compared to Tango-77 wild type protein. In addition, the anti-Tango-77

antibodies of the invention can be used to detect and isolate Tango-77 proteins and modulate Tango-77 activity.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to Tango-77 proteins or have a stimulatory or inhibitory effect on, for example, Tango-77 expression or Tango-77 activity.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent Nos. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In another embodiment, an assay is used to determine the ability of the test compound to modulate the activity of Tango-77 or a biologically active portion thereof, for example, by determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule. As used herein, a "target molecule" is a molecule with which a Tango-77 protein binds or interacts in nature, for example, a molecule on the surface of a cell. A Tango-77 target molecule can be a non-Tango-77 molecule or a Tango-77 protein or polypeptide of the present invention. In one embodiment, a Tango-77 target molecule is a component of a signal transduction pathway, for example, Tango-77 may bind to a IL-1 receptor or another receptor thereby blocking the receptor and inhibiting future signal transduction. Determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule can be accomplished by one of the methods described above. In a preferred embodiment, determining the ability of the Tango-77 protein to bind to or interact with a Tango-77 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a Tango-77-responsive regulatory element operably linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, inflammation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a Tango-77 protein or biologically active portion thereof with a test compound and determining the ability of the

test compound to bind to the Tango-77 protein or biologically active portion thereof. Binding of the test compound to the Tango-77 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the Tango-77 protein or biologically active portion thereof with a known compound which binds Tango-77 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a Tango-77 protein, wherein determining the ability of the test compound to interact with a Tango-77 protein comprises determining the ability of the test compound to preferentially bind to Tango-77 or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting Tango-77 protein or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the Tango-77 protein or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of Tango-77 can be accomplished, for example, by determining the ability of the Tango-77 protein to bind to a Tango-77 target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of Tango-77 can be accomplished by determining the ability of the Tango-77 protein to further modulate a Tango-77 target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting the Tango-77 protein or biologically

active portion thereof with a known compound which binds Tango-77 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a Tango-77 protein, wherein determining the ability of the test compound to interact with a Tango-77 protein comprises determining the ability of the Tango-77 protein to preferentially bind to or modulate the activity of a Tango-77 target molecule.

It is possible that membrane-bound forms of Tango-77 exist. The cell-free assays of the present invention are amenable to use of both the forms Tango-77. In the case of cell-free assays comprising a membrane-bound form of Tango-77, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of Tango-77 is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either Tango-77 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to Tango-77, or interaction of Tango-77 with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels

include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ Tango-77 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical Co.; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or Tango-77 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of Tango-77 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either Tango-77 or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated Tango-77 or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with Tango-77 or target molecules but which do not interfere with binding of the Tango-77 protein to its target molecule can be derivatized to the wells of the plate, and unbound target or Tango-77 trapped in the wells by antibody conjugation. Methods

for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the Tango-77 or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the Tango-77 or target molecule.

In another embodiment, modulators of Tango-77 expression are identified in a method in which a cell is contacted with a candidate compound and the expression of Tango-77 mRNA or protein in the cell is determined. The level of expression of Tango-77 mRNA or protein in the presence of the candidate compound is compared to the level of expression of Tango-77 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of Tango-77 expression based on this comparison. For example, when expression of Tango-77 mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of Tango-77 mRNA or protein expression. Alternatively, when expression of Tango-77 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of Tango-77 mRNA or protein expression. The level of Tango-77 mRNA or protein expression in the cells can be determined by methods described herein for detecting Tango-77 mRNA or protein.

In yet another aspect of the invention, the Tango-77 proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi

et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with Tango-77 ("Tango-77-binding proteins" or "Tango-77-bp") and modulate Tango-77 activity. Such Tango-77-binding proteins are also likely to be involved in the propagation of signals by the Tango-77 proteins as, for example, upstream or downstream elements of the Tango-77 pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for Tango-77 is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming an Tango-77-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with Tango-77.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequence identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, the sequence can be used to: (i) map the respective gene on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, Tango-77 nucleic acid molecules described herein or fragments thereof, can be used to map the location of the Tango-77 gene(s) on a chromosome. The mapping of the Tango-77 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, a Tango-77 gene can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the Tango-77 sequences. Computer analysis of Tango-77 sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the Tango-77 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and

mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow (because they lack a particular enzyme) but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the Tango-77 sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a Tango-77 sequence to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical, e.g., colcemid that disrupts the mitotic spindle. The

chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al. (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease

associated with the Tango-77 gene can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The Tango-77 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the Tango-77 sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the

sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The Tango-77 sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from Tango-77 sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Tango-77 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the Tango-77 sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 or 30 bases.

The Tango-77 sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for

example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such Tango-77 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., Tango-77 primers or probes can be used to screen tissue culture for contamination (i.e., screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining Tango-77 protein and/or nucleic acid expression as well as Tango-77 activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant Tango-77 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with Tango-77 protein, nucleic acid expression or activity. For example, mutations in a Tango-77 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with Tango-77 protein, nucleic acid expression or activity.

Another aspect of the invention provides methods for determining Tango-77 protein, nucleic acid expression or Tango-77 activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of Tango-77 in clinical trials.

These and other agents are described in further detail in the following sections.

1. <u>Diagnostic Assays</u>

An exemplary method for detecting the presence or absence of Tango-77 in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting Tango-77 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes Tango-77 protein such that the presence of Tango-77 is detected in the biological sample. A preferred agent for detecting Tango-77 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to Tango-77 mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length Tango-77 nucleic acid, such as the nucleic acid of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent

conditions to Tango-77 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting Tango-77 protein is an antibody capable of binding to Tango-77 protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect Tango-77 mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of Tango-77 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of Tango-77 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of Tango-77 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of Tango-77 protein include introducing into a subject a labeled anti-Tango-77 antibody. For example, the antibody can be

labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting Tango-77 protein, mRNA, or genomic DNA, such that the presence of Tango-77 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of Tango-77 protein, mRNA or genomic DNA in the control sample with the presence of Tango-77 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of Tango-77 in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of Tango-77 (e.g., an immunological disorder). For example, the kit can comprise a labeled compound or agent capable of detecting Tango-77 protein or mRNA in a biological sample and means for determining the amount of Tango-77 in the sample (e.g., an anti-Tango-77 antibody or an oligonucleotide probe which binds to DNA encoding Tango-77, e.g., SEQ ID NO:1 or SEQ ID NO:3 or SEQ ID NO:6, or SEQ ID NO:10). Kits may also include instruction for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of Tango-77 if the

amount of Tango-77 protein or mRNA is above or below a normal level.

For antibody-based kits, the kit may comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to Tango-77 protein; and, optionally (2) a second, different antibody which binds to Tango-77 protein or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit may comprise, for example: (1) an oligonucleotide, e.g., a detectably labelled oligonucleotide, which hybridizes to a Tango-77 nucleic acid sequence or (2) a pair of primers useful for amplifying a Tango-77 nucleic acid molecule;

The kit may also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit may also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit may also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of Tango-77.

2. <u>Prognostic Assays</u>

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant Tango-77 expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or

at risk of developing a disorder associated with aberrant expression or activity. Thus, the present invention provides a method in which a test sample is obtained from a subject and Tango-77 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of Tango-77 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant Tango-77 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant Tango-77 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease Tango-77 activity). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant Tango-77 expression or activity in which a test sample is obtained and Tango-77 protein or nucleic acid is detected (e.g., wherein the presence of Tango-77 protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant Tango-77 expression or activity).

The methods of the invention can also be used to detect genetic lesions or mutations in a Tango-77 gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant

inflammation. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding a Tango-77-protein, or the mis-expression of the Tango-77 gene. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from a Tango-77 gene; 2) an addition of one or more nucleotides to a Tango-77 gene; 3) a substitution of one or more nucleotides of a Tango-77 gene; 4) a chromosomal rearrangement of a Tango-77 gene; 5) an alteration in the level of a messenger RNA transcript of a Tango-77 gene; 6) an aberrant modification of a Tango-77 gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a Tango-77 gene; 8) a non-wild type level of a Tango-77-protein; 9) an allelic loss of a Tango-77 gene, and 10) an inappropriate post-translational modification of a Tango-77-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions or mutations in a Tango-77 gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the Tango-77-gene (*see,*

e.g., Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a Tango-77 gene under conditions such that hybridization and amplification of the Tango-77-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a Tango-77 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA

indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see*, e.g., U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in Tango-77 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations in Tango-77 can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al. supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the Tango-77 gene and detect mutations by comparing the sequence of the sample Tango-77 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a

variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see, e.g.*, PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the Tango-77 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type Tango-77 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called

"DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in Tango-77 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a Tango-77 sequence, e.g., a wild-type Tango-77 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in Tango-77 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control Tango-77 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in

electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic*

Acids Res. 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a Tango-77 gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which Tango-77 is expressed may be utilized in the prognostic assays described herein.

3. <u>Pharmacogenomics</u>

Agents, or modulators which have a stimulatory or inhibitory effect on Tango-77 activity (e.g., Tango-77 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g., acute or chronic inflammation and asthma) associated with aberrant Tango-77 activity. In

conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of Tango-77 protein, expression of Tango-77 nucleic acid, or mutation content of Tango-77 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.,* Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-

malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM shows no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of Tango-77 protein, expression of Tango-77 nucleic acid, or mutation content of Tango-77 genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes

to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a Tango-77 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of Tango-77 (e.g., the ability to modulate aberrant inflammation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase Tango-77 gene expression, increase protein levels, or upregulate Tango-77 activity, can be monitored in clinical trials of subjects exhibiting decreased Tango-77 gene expression, decreased protein levels, or downregulated Tango-77 activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease Tango-77 gene expression, decrease protein levels, or downregulate Tango-77 activity, can be monitored in clinical trials of subjects exhibiting increased Tango-77 gene expression, increased protein levels, or upregulated Tango-77 activity.

For example, and not by way of limitation, genes, including Tango-77, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates Tango-77 activity (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared

and analyzed for the levels of expression of Tango-77 and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of Tango-77 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a Tango-77 protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the Tango-77 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the Tango-77 protein, mRNA, or genomic DNA in the pre-administration sample with the Tango-77 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of Tango-77 to higher levels than detected,

i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of Tango-77 to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) developing or having a disorder associated with aberrant Tango-77 expression or activity. Alternatively, disorders associated with aberrant IL-1 production can be treated with Tango-77. Such disorders include acute and chronic inflammation, asthma, some classes of arthritis, autoimmune diabetes, systemic lupus erythematosus and inflammatory bowel disease.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant Tango-77 expression or activity (or aberrant IL-1 expression or activity), by administering to the subject an agent which modulates Tango-77 expression or at least one Tango-77 activity. Subjects at risk for a disease which is caused or contributed to by aberrant Tango-77 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the Tango-77 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of Tango-77 aberrancy, for example, a Tango-77 agonist or Tango-77 antagonist agent can be used for treating the

subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating Tango-77 expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of Tango-77 protein activity associated with the cell. An agent that modulates Tango-77 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a Tango-77 protein, a peptide, a Tango-77 peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of Tango-77 protein. Examples of such stimulatory agents include active Tango-77 protein and a nucleic acid molecule encoding Tango-77 that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of Tango-77 protein. Examples of such inhibitory agents include antisense Tango-77 nucleic acid molecules and anti-Tango-77 antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a Tango-77 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) Tango-77 expression or activity. In another embodiment, the method involves

administering a Tango-77 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant Tango-77 expression or activity.

Stimulation of Tango-77 activity is desirable in situations in which Tango-77 is abnormally downregulated and/or in which increased Tango-77 activity is likely to have a beneficial effect. Conversely, inhibition of Tango-77 activity is desirable in situations in which Tango-77 is abnormally upregulated and/or in which decreased Tango-77 activity is likely to have a beneficial effect.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLES

Example 1: Isolation and Characterization of Human Tango-77 cDNAs

Cytokine genes IL-1α, IL-1β and IL-1ra have been found to be closely clustered on chromosome 2, i.e., IL-1α, IL-1β and IL-1ra are located within 450 kb of each other. BAC clones containing IL-1α and IL-1β were used to identify other proximal unknown cytokine genes. To do this, a BAC clone containing IL-1α and IL-1β was selected from a BAC library (Research Genetics, Huntsville, Alabama) using specific primers designed against IL-1α and IL-1β. The DNA from the BAC was extracted and used to make a random-sheared genomic library. From this BAC library, 4000 clones were selected for sequencing. The resulting genomic sequences were then assembled into contigs and used to screen proprietary and public data bases. One genomic contig was found to contain two

segments of sequences which resemble IL-1ra. These two segments are potential exons of Tango-77 gene.

Two PCR primers were then designed from the two potential exons and used to screen a panel of cDNA libraries for the expression of a Tango-77 message. A cDNA library from TNF-α treated human lung epithelia showed a positive band of the predicted size (i.e., if the two exons are spliced together). Using the PCR fragment as a probe, a single cDNA clone was isolated from the same library. This cDNA contains an insert of 989 bp. The cDNA clone contains three possible open reading frames. The first open reading frame encompasses 534 nucleotides (nucleotides 356-889 of SEQ ID NO:1; SEQ ID NO:3) and encodes a 178 amino acid protein (SEQ ID NO:2). This protein may include a predicted signal sequence of about 63 amino acids (from amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID NO:4)) and a predicted mature protein of about 115 amino acids (from about amino acid 64 to amino acid 178 of SEQ ID NO:2 (SEQ ID NO:5)).

The second putative nucleotide open reading frame encompasses 498 nucleotides (nucleotides 389-889 of SEQ ID NO:1; SEQ ID NO:6) and encodes a 167 amino acid protein (SEQ ID NO:7). This protein includes a predicted signal sequence of about 52 amino acids (from amino acid 1 to about amino acid 52 of SEQ ID NO:7 (SEQ ID NO:8)) and a predicted mature protein of about 115 amino acids (from about amino acid 53 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:9)).

The third open reading frame (nucleotides 372-889 of SEQ ID NO:1; SEQ ID NO:10) encompasses 408 nucleotides and encodes a 136 amino acid protein (SEQ ID NO:11). This protein includes a predicted signal sequence of about 21 amino acids (from amino acid 1 to about amino acid 21 of SEQ ID NO:11 (SEQ ID NO:12)) and a predicted

mature protein of about 115 amino acids (from about amino acid 22 to amino acid 136 of SEQ ID NO:11 (SEQ ID NO:13)).

Tango-77 is predicted to be 35% identical to human IL-1ra at the amino acid level.

Example 2: Expression of Tango-77 mRNA in Human Tissues

The expression of Tango-77 was analyzed using Northern blot hybridization. A PCR generated 989 bp Tango-77 product was radioactively labeled with ^{32}P-dCTP using the Prime-It kit (Stratagene; La Jolla, CA) according to the instructions of the supplier. Filters containing human mRNA (MTNI and MTNII: Clontech; Palo Alto, CA) were probed in ExpressHyb hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Tango-77 mRNA was not detected in any unstimulated tissues (brain, liver, spleen, skeletal muscle, testis, pancreas, heart, kidney and peripheral blood leukocytes) mRNA on Clontech Northern blots.

Over 96 cDNA libraries were then tested for the presence of Tango-77 using PCR amplification. Only three libraries displayed a positive signal. These libraries were the TNFα-treated bronchoepithelium, TNFα-treated SSC cell line and anti-CD3-treated T cells.

Example 3: Characterization of Tango-77 Proteins

In this example, the predicted amino acid sequence of human Tango-77 protein was compared to the amino acid sequence of known protein IL-1ra. In addition, the molecular weight of the human Tango-77 proteins was predicted.

The human Tango-77 cDNA (Figure 1; SEQ ID NO:1) isolated as described above encodes a 178 amino acid protein (Figure 1; SEQ ID NO:2) or a 167 amino acid

protein (Figure 1; SEQ ID NO:7) or a 136 amino acid protein (Figure 1; SEQ ID NO:11). The signal peptide prediction program SIGNALP Optimized Tool (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that Tango-77 includes a 63 amino acid signal peptide (amino acid 1 to about amino acid 63 of SEQ ID NO:2 (SEQ ID NO:4)) preceding the 115 mature protein; or preceding the 115 mature protein (about amino acid 52 to amino acid 167 of SEQ ID NO:7 (SEQ ID NO:8)); or preceding the 115 mature protein (about amino acid 21 to amino acid 136 of SEQ ID NO:11;SEQ ID NO:12).

As shown in Figure 2, Tango-77 has a region of homology to IL-1ra (SEQ ID NO:14).

Mature Tango-77 has a predicted MW of about 13 kDa and the predicted MW for the immature Tango-77 is 19.6 kDa, 18.5 kDa or 15.2 kDa, not including post-translational modifications.

Example 4: Preparation of Tango-77 Proteins

Recombinant Tango-77 can be produced in a variety of expression systems. For example, the mature Tango-77 peptide can be expressed as a recombinant glutathione-S-transferase (GST) fusion protein in *E. coli* and the fusion protein can be isolated and characterized. Specifically, as described above, Tango-77 can be fused to GST and this fusion protein can be expressed in *E. coli* strain PEB199. Expression of the GST-Tango-77 fusion protein in PEB199 can be induced with IPTG. The recombinant fusion protein can be purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads.

Example 5: Alternatively spliced forms of IL-1ra and Tango-77

Computer program Procrustes (Gelfand et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:9061-9066) is an alignment algorithm that predicts the presence of alternatively spliced exons for a protein of interest in a stretch of genomic DNA. Using the IL-1ra sequence, Proscustes was used to search for the presence of additional sequences that might encode for alternatively spliced forms of IL-1ra in the two overlapping BAC genomic sequences (see Fig. 3 and Fig. 4). Potential sequences that encode variant exons for IL-1ra were identified. These predicted exons aligned well with the N-terminal region of IL-1ra, but were not present in Tango-77. The results from Procrustes predicts the existence of more spliced forms of IL-1ra.

Furthermore, Procrustes also predicted an additional sequence in BAC1 and BAC2 that encodes an alternatively spliced exon for Tango-77 (T77-procrustes; Fig. 5). This predicted splice variant form of Tango-77, T77-procrustes, was aligned with Tango-77 (Fig. 6) and with IL-1ra and IL-1β (Fig.7).

PCR primers within this sequence can be used to generate a product that can be used to screen a panel of cDNA libraries using standard techniques. Suitable cDNA libraries include libraries made from TNFα-treated bronchoepithelium, TNFα-treated SSC cell line and anti-CD3-treated T cells. The resulting cDNA clone(s) can be isolated from the library and sequenced to identify additional Tango-77 cDNAs.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 45% identical to the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof;

c) nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or the polypeptide encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9,

SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10, or the complement thereof under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10 or the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell containing the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, or SEQ ID NO:13.

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, SEQ ID NO:10 or the complement thereof under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807;

b) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, or an amino acid sequence encoded by the cDNA insert of the plasmid deposited with ATCC as Accession Number 98807, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid sequence of

SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:6, or SEQ ID NO:10 under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for Tango-77-mediated signal transduction.s

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

GTCGACCCACGCGTCCGCAGACGTCTACCTGGGGGTCCCGTCTGCGCTCCCGGGATGGAAAACGCCCAGGGGAAACTTA 79
GGCAGGCGAGCGGACGGGCACCTCCCGCGGGACGAACTCACTCGGTGGCCTCCTACTTCCCCGGCCGTGTTCCAACGCC 158
TGAGAATAACGGGAACAGCGGTCGTACTCACCGACAGCGGCAGCAGCGGCCTCTCTCAATTGGGCAAAGCACTCCAGAC 237
TTTTTGGAAGAGTGACACCAAAGGCAAGCACCTGCTTGGCAGGCCCCTCAGCTTCTACGCAAGTATAAGTCTTGGACTT 316

M S F V G E N S G V 10
CATTCCATTTTCTGTTGAGTAATAAACTCAACGTTGAAA ATG TCC TTT GTG GGG GAG AAC TCA GGA GTG 385

K M G S E D W E K D E P Q C C L E D P A 30
AAA ATG GGC TCT GAG GAC TGG GAA AAA GAT GAA CCC CAG TGC TGC TTA GAA GAC CCG GCT 445

G S P L E P G P S L P T M N F V H T K I 50
GGA AGC CCC CTG GAA CCA GGC CCA AGC CTC CCC ACC ATG AAT TTT GTT CAC ACA AAG ATC 505

F F A L A S S L S S A S A E K G S P I L 70
TTC TTT GCA TTA GCC TCA TCC TTG AGC TCA GCC TCT GCG GAG AAA GGA AGT CCG ATT CTC 565

L G V S K G E F C L Y C D K D K G Q S H 90
CTG GGG GTC TCT AAA GGG GAG TTT TGT CTC TAC TGT GAC AAG GAT AAA GGA CAA AGT CAT 625

P S L Q L K K E K L M K L A A Q K E S A 110
CCA TCC CTT CAG CTG AAG AAG GAG AAA CTG ATG AAG CTG GCT GCC CAA AAG GAA TCA GCA 685

R R P F I F Y R A Q V G S W N M L E S A 130
CGC CGG CCC TTC ATC TTT TAT AGG GCT CAG GTG GGC TCC TGG AAC ATG CTG GAG TCG GCG 745

A H P G W F I C T S C N C N E P V G V T 150
GCT CAC CCC GGA TGG TTC ATC TGC ACC TCC TGC AAT TGT AAT GAG CCT GTT GGG GTG ACA 805

D K F E N R K H I E F S F Q P V C K A E 170
GAT AAA TTT GAG AAC AGG AAA CAC ATT GAA TTT TCA TTT CAA CCA GTT TGC AAA GCT GAA 865

M S P S E V S D * 179
ATG AGC CCC AGT GAG GTC AGC GAT TAG 892

GAAACTGCCCCATTGAACGCCTTCCTCGCTAATTTGAACTAATTGTATAAAAACACCAAACCTGCTCACTAAAAAAAAA 971
AAAAAAAAGGGCGGCCGC 989

Fig. 1

```
                1                                                       50
IL1ra-human   MEICRGLRSH LITLLLFLFH SETICRPSGR KSSKMQAFRI WDVNQKTFYL
  T77-human   ---------- ---------- ---------- ---------- ----------
 IL1b-human   ---------- ---------- ----APVRSL NCTLRDSQQK SLVMSGPYEL
  Consensus   ---------- ---------- ---------- ---------- ----------

              51                                                     100
IL1ra-human   RNNQLVAGYL QGPNVNLEEK IDVVPIEPH. ALFLGIHGGK MCLSCVKSGD
  T77-human   ----MNFVHT KIFFALASSL SSASAEKGS. PILLGVSKGE FCLYCDKDKG
 IL1b-human   KALHLQGQDM EQQVVFSMSF VQGEESNDKI PVALGLKEKN LYLSCVLKDD
  Consensus   ---------- ---------- ---------- ---LG----- --L-C-----

              101                                                    150
IL1ra-human   ETR..LQLEA VNITDLSENR KQDKR.FAFI RSDSGPTTSF ESAACPGWFL
  T77-human   QSHPSLQLKK EKLMKLAAQK ESARRPFIFY RAQVGSWNML ESAAHPEWFI
 IL1b-human   K..PTLQLES VDPKNYP..K KKMEKRFVFN KIEINNKLEF ESAQFPNWYI
  Consensus   -----LQL-- ---------- ------F-F- ---------- ESA--P-W--

              151                                        192
IL1ra-human   CTAMEADQPV SLTNMPDEGV MVTKFYFQED E--------- --
  T77-human   CTSCNCNEPV GVTDKFENRK HI.EFSFQPV CKAEMSPSEV SD
 IL1b-human   STSQAENMPV FLGGT.KGGQ DITDFTMQFV SS-------- --
  Consensus   -T------PV ---------- ----F--Q-- ---------- --
```

FIG. 2

2/118

>Contig1
GAAGTGAAGATATAATGTATAGTAGTAATATATAATGTTAGGTGAATTAA
AGGAAATAGAATATATTGGGGAGTAATTATGGGTGTAAAGAAATATAGTA
GGGAAGTATTTAGATTTGAGAAAAAAAAAAAGGAATTTAGTGTAGGTGAA
NAATAAAAGNANAAGGTTAAAAATTAAAAAAAAATTAAATATAAATAAAT
AAATAAAAATAAAAATAAAATAAAAAATTTAAAAAATTAAAAAAATATAA
AAAATAAAGAAATGGAAGTGGATTCTTAGAAAAAAAAGAAAGTAAGGTGA
TATGAGGAGATAGAGAGGATGTGGTGTGAGATGATTGGTTTAATTAGAAA
ATAGGTTTTGAATAGAGTGGGAAAGTAGAGTTTTGGTAAATGTGGGGGGA
AGAGGGTAATGTTGTTTGAGTGAAAGAAAAAATGGTATATTTTTATAAAA
TAATGAGGAAAGTGTGTGAAAAAAAAATTATTGGGATTTGGGAAGGTGAT
ATATAAAGTTGTGGAAAATTTGGGGGGTGGGGTTTATTTAGGATTAAAAA
GTTATTTAAAGAATGAAAATGAATTTTTGTTTGTAATTTGGGGATAAGAA
ATTAATGTTTAGAAAGAAAGGGAAAAAATTGAAGAAAAAAATTTAGATTT
TGGAAATTTAAAAATATTGTGGGTGTAAATAGGAAGGATTTTTAAAGGTA
ATTGTGGAAGGGATTTGTGTGGAAAATAATAGGGAGAAAAAATGGGG
>Contig2
GCATCTAACTGGAGCCTGCATTATTACAGATTTAGCATCACCAAAGTCTA
AACAATTAGACTGACTAAGGCAGAACTGCCCTTATGACAGCAGACATAAG
AAGGAAAAGGCCAAAACACTGTGTTAAAAATTATCCAAATGTGAGGAAAA
GGCAAAGAGAGTAGGTGTGCCTTTTTAGTGTCTAAGCTGCCTGCCCAAGG
GGCATCTGATGCTCTCAGGCAGGAGTCCACAAATTTTTTTTTGTAAAAGA
TCAGATAGTAAATCTTTTCAGCGTGAAGAGCATGAGGTCTCTGTCACAAA
TACTCAACCACCATTACAACATGAAAGCAGCCAACAGACAACACATGACA
AATGAGTGTGGCTGTGTTCCAGTAAATCTTGATTACAAAAACAGGCAAGA
GGCCAGAGCTGACCCATGGGCCATAGTTTGCTGACCCCTTCTGTAAAGGA
AAGTATTTTTGTTTGACTTGCTGTTTACCATTGATTGAACACAAGGCTCT
GTAAAGTTACTTGTTAACTTGCAGAAGATTGATGAGTGGCAAGTAATTTT
TATTCACCAGAATATAAAATTATTTCTGTTCAGTAGAAAAGATAAACCAA
CTGTGATATTATGGTCCTG
>Contig3
GGGGTGTCTGTCTACCATGTGCTCGCAGTTCTGTAATAAATGTTCTCTCA
AGATCCTTAAAATCTCTTGGAAATTATAAAAATATTGGAAAGAGAAGAAC
AGTTTTTAAAATATATATATATATATATATATTTTTTTTGAGATGGAGTCTT
GCTCTGTCGTCCAGGCTGGAGTGCAGTGGCGCAAACTTGGTTCACCACAA
CCTCTGCCTCCCGGGTTCAAGCGATTCTTCTGCCTCAGCCTCCTGAGTAG
CTGGGACTACAGGCGCCCGCCACCACGCCCAGCTAATTTTTGTATTTTTA
GTAGAGACGAGGTTTTACTATGTTGGCTAGGCTGGTCTCAAACTCCTGAC
CTTGTGATCTGCCCGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTG
AGCCACTGCACCTGGCCAGTTTTTTAAATATATTTTTAAAAACACTTGAA
TAAGAGTCAGTGTAAACTAGAAGTTTAAAAATGCTTCACAGAACACCCAG
GGTTTACATTACAAGATTCTCACAACAAACCTATTGTAAAGGTGAGTAAG
GCATGTTATTACAGAGAAAAGTTTGGGAGCAAAACTGTAAAAAATTATAT
TTTTGTTGTATTTTCTAAGAGAAAGAGTATTGTTATGTTCTCCTAACCTC
TGTTGATTACTACTTTAAGTGATTTCCTTGAGAGCACATGATGATCC
>Contig4
GCCGTTCATAGAAAACTGAAAGCAATAAGATGACTAGGTAAGCATGACAT
TTAAAAGGTATTCATGGGACGTGGTTACAAAACCAACTCACAACTAAAAA
GTCTTAGGACCTCTCGCTGACTTAGGAGCCTGATCCCAACTCTGAGAATG
ACTCAGTGTGTTACCCTGTGGCTAGTGTAGACCAATGATCCTGTCTCAGA
GTCACTAGCCAACAGCCCATATCAAGTACTTGAAACTTTGACTCAGAAAC
CTCAGTGTCAGAACCTTTGACCTAGGAACCACCTGTAGTGGTTAACTGCA
ATTTGCACCCCTTAGTTCAGGGCTTTACAACACCGGGGGCGGGGAGGGGA
AAGGCATANANCTGATGACCTAAAGGAAACCCATTGCAGCAACGCTTTTG
TGTTAAGTGTACAAATAAGTGTTGTTTTAGAATCCTCCAGGTAATGCCTT
TGTTATTTAATGTGTCTGAGACAATTCTGCACATTAAAGAATATAAAATA
TTACCTTGTAATTCCAATTTGAAATGTGTAATTGACATTAGACTTCTATT
TGAATTTGAAATGTCTAAAACAATGTGGTTAAGTTTGTAAAAGGTGTGTG
AATTTTGAGTCTGATTTACTACATTTTTTTTTAATTTTCTTTTTTTTTGG
AGTTTTAGGGATTGCTTAGATGGCTAGAAAGATTTTATTCATCAGATTTT

FIG. 3 (1 of 52)

TAAGTCTGCCTTGGCAGGCACTTGCAG..TTTGAAAGAATCAGATATATC
AAATTTGTAGTTTAAAATATTTAAGGGAACTCAATTAACTATGCTAGAAA
AGAGAATTAAGTATTTAGGAGGATTTAATATGGTGTGAAAGTTGTGAAAA
TCAAAATGGAGACACTAATGTTAAGAAAACCCTGATAAATGGAACCAGGG
AAAGGCATGAAGATAGAGTTCTCACACTTGTATCCCTGATCATGAAAAAG
ATCTGC
>Contig5
GGGTTTTTCCGCGTTTTTACCCGAAATCTTCAAGGGATGGGAAAAAGAAA
ATTGCTAAAAAATCTCGGTTTTTTGGTTTTAACAGATATTTACACCNTGG
ATCCCATTTATTATGTTGTCCCCAAGGTTTTCGGTGGGTTCCCAATCAGT
TAGCCCCCCTCCACAGTGAAAGCACTTTACTTTATCACCTTCACCTAAAG
CATAAAATCCAGCTCTTGAAAGCTGCTCCTTGTTAACTGAATATATCCAC
ATCCCAAAAGTAATGATCCATGCTTCATAATCTGCCACGGATGGATGGAT
GGATGGATGGATGGATGGATGGATGAATGGATGGATTGATTTCTTG
GAGGATTTGTTGAATTTGGGAAATTCCACGCCAGGACAGCTGGCCCAAAC
TGCCCGCGACAATCTGCTCGGTACAAGGGGAGGGTCCTGGAGAGGGTGCG
GCCCGAGCCCCAGTTTGGAAATGCCAACTTGGCTCTGCAGCCGGGCCTTA
GCCACTTGGGTCTGGCGTCCCTCCATTATTAGCGCCATGCCGGCTCGGGG
TGCTGCCAAGTCCCTGAGAGCACAAGCC
>Contig6
CGCGCTCAAGAAAAGCTGAAGTGTGAATGTTCTGTCTACCTTCACAGTAA
ATGCTAAGAGAATGACCCAAGAGCAGAGGGTATCACTCTGCTACGGAGGA
TTGATTGTAACTGGCTCTCCTGCCTTAGCAAGAAATGCCAGAACCATGGT
CATTCAAGTTCTTGACCAAAAACTGCCTTCATGAGAATCAACTTCCCCAA
GAAAAAAAAAGCAGAAACAGGCAAAGCTTCCAGCATGGTAGGTAATACTG
ACCCTTCTTCCCTCCTTCCTTTGGAGATTCACACAGTAATAATGCATAAA
GCTTTGCCAATGGACTAAGCACTGCCCAGGGGTTTTTGTCATGCCTGGAC
TGAAATGCTCTTTTTGCGTTATCATAGAATCCCAGTGCAGTCTGAGTAGA
CTCTAAGCAAAAGGGACATTTTTCAAAAAGGCTTTAAATTGCTAGTACAA
AGAAGGCAACAAAACTTGCGTAACTGTGGACAGATTAACTCACTTGGTGT
TTTGGCTCTTCAGTTTTCCCTTGGCTGCGAAGTACTCCTGAAGCTTTCTC
TGCGGCTCTTCCTGCAAGCAGGCAAGCAAAAAAACGACTGAACTTTATTT
CGAGAT
>Contig7
GAAGAGCCGCTAACTTGCTGTAGTGATAAGGAATGAACTAAGGCTAGGGA
CATATTAACATCCGCTGGTGGTGACTCTTTAGCCTAGATCTTACCCCACT
CCTGCTCCTTCCATATGGTTCGGTCTCAGGCTCACTACCGATCAATGGCG
TACTAAAAGCACTAACTATAGACTCCAACACGTCTGTCGTGTGTTTCACG
ACAAGCCGTGGAGTTAATCCCTCTGACAGTAGCTCAGATAAGGATGGGCT
ATCATGGGCCCGGAACTGGGGCATGACGCTCGTCACCAACGCATGAGCTC
CCCAAGTATGCTATACCTGTCCCTATGAAGGGCTTCCAACTCTATGTGCA
GTCCCCATGTGGAGAGTCAGGTATTGATTGATCAAGCCAGGGGTGTGGTG
AATGGGGAGCTTCCTACAGGGGTAATGATAATTGAAATGCACGGTGATGG
GGATTTTCATATTGGTCTCCTAAGGAGATAACAGATTGGATGCGGGGTCG
ATATTCCACTGCCCAGGGTGTGTACCGAGGGTATCTGCAGGTGGATCTCC
TCCCCACGTTTGATTAATACTCCTGTCTTGGGAAGCATAGACGGGCGGGG
GAAATGATGAAGGGTGACCACTCCCC
>Contig8
GGGAACGCAGTGCTCTGTACGATGGCCTTGATTGCGAATTCCTGCAGGGG
GGG
>Contig9
GGCAAGAGATTTAATATTCATTCCATCTTCATTTGGAAGATGAAAAATTG
GGGACCAGAGAGGGGAGGGGACTGGGCCAAGTTTTCAAAGAAAAGTCAGT
AGGAATTGTGAATTCCTGGGGGCCGGGGCCCATTAGTGCTGTTTTGGATC
AGTAAATGGAGATGTGAGTTTCAACAGTAACAGGGACATTTTAAAATTAA
AATGATTTAACCTTTAGAAAATGTCCTATTTTGTAATAATGATGGATTCA
CAGGAAGGTACAAAGAAATGTCCAGAGAGTTCNTGAGCCCCCTTCAGCCA
GCTTCTTCCAATGTTAACATCTTGCATTATTATAGTACAACATCAAAACT
GGGAAATCGATATTGGTACTGTCCAGATAGCTTACTCAGATTTTGCCAGT
TATACTTCCACTCATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG

FIG. 3 (2 of 52)

TGTGTGTAGCTCTATGCAATTTTATGT..GTAGCTTCATGTAACCACCAC
AATCACAATACTTAACTATGCCCTCATCACAAGACTCTCTCTTGCTATGC
TTTACAGCTGTATCCTCTTCATCTCCAAACCCTAAGCCCACCTCACCGCC
TCCACCATCTCTAATCCCTGGCAACCACTATTCTGTGCTCCATCTCTGTA
ATTAATTGTGTTAATTAATGTTATACAAATGGAATCATGAAGTATGTGTC
CTTTGAGATTGGGCTGTTAATTTTTCACTCAGCACAATTTCCGTGAGTCT
AATCCAACTTGTGTGTAGCAGTAATTCTTTCCTTATTATTGCTGAATAAT
ATGCCATGGTATGGATGTATCACAGTGTGTCTAATCCTTTGCCCATTGAA
AGGAATTTGGATAATTTCCAGGTTTTGGCTATTATGAATAAAGTGAACAT
AAGACATGTGTGTACAAATTTTGGTGTGATCAAAAGTCTCATTTCTCTGG
GATAAATGCCCGGTAATGAAATGGCTGGGTTGTGTGGG

>Contig10
GCAAGAACACAGGCGCGTATTATAACCTTACTACCAAGACCTGAACCCAT
ATAAAGGTTTATGCGTAACAATCATCATCCCTGTTCCAGAAGATTACACG
TACGACCACGCCTGGCTCACCGACTCACGTGGGCCAGTACCAGAAATTCT
CCCAAACAAACAGTCGTGTCTGAAAACAATCGCGGTGACCTCCACGGTTA
GAAAAGCCTGTTTTCAAGTCCTGGAATTGCCACATATTAGCTGGGTAACT
TTGGGCATCACATTTACTCTCTCCGAATTTCAGATTGCAAAAACTCATTG
GATTGTTTTGTGGATTGAAAGAAATAATGTAAATTTAGGCCGAGTGCTTT
GACTTACGCCTGTAATCCTATCACTTTGGGAGGCCAAAGCAGGAGGGTCA
CTTGAGCTCAGGAATTTGAGACCACCTCTGGCAACATAGTGAGATCCTGT
CTCTACAAAAAATTTTTTTAAATTATCCAGCATGGTGGTACACGCCTGT
ATTCCCAGCTACTCAGGAGACTGAGGTGTGAGGATTGCTAGAACCTGGGA
GATCAAGTCAACAGTGAGCCGTGGTTGTGCCACTGCCCTCCAACCTCAGT
GACAGAGGAAGACCCTGTCTCAAAAAAAAAAAAAAAAGTAGTAAGTTTAA
AGAACTTAGTGTAGGCCTGGCATATAAATGATATTGTTGATGTTGATGTT
AGCTTGAAGGCACATTTATAGGAGTAGGGATTTTATAACATTATGAGCCT
GAGAGCACATATAATGTTCCC

>Contig11
GGTCTAACATGCTCCAACTGAAGAAACCCCACACTTGTCCGGCAAGGAAA
CTACTACAGATTTCCTGACCTACTGTGCAATTCGGGGCATGCGACGGGAC
TGTGTTTCTGGGTACGCTGTCTCAGGTTCGTCTGGGATGTAAGAATTCAA
CTTCAGTAGTTCTCTCATAGACGCCGACGAGAGGGGCGTCTCTTTTCTCT
GATGAATCTGCCAGATCTTCCACTTCATAGAGTCTAAATCCTCCGATTCG
ATCTACTGGAGACCCCCACGTTACAAAAACGTCTAACGTCGGTGACAGCT
CCCCACATAGGGAAAGATCACCTGAGTCTCACTACCTCACATTAGTGCTA
TCTCCAGCCCCATGCTATCTACGAGATGGTCACGCGAGGTTTAAGGGGTC
TCCGATTCCGGTGGTCCGATTCAGCTAATCGTGGCCCTACGTGAACGATC
ACTCCTGCTCGTAACATCGATACAGGGTCGCGCTGACAAATGGTACTACG
TAGGTTCTCAGGTCAATGCCGCGTCACGAATGAGCCTAACTACCCCATAA
GTGCACGTACTGTGTTACCTTTCCTGTTCGGCCAAACCTGCTACTGTATG
CTGTGCTTGTTT

>Contig12
AGGCTCCATGTGCTCTAGCCTGATTATCTTTTCAAGTGTTTTATTTGCTA
ATCTATAAGGCCCCTTTCGTAAAATGTTCACTCATTTTCTAATTAGATAT
TTTTTTTAATGTTGAGTTTTGAGAGTTCTTTAGATATTTTAGATACAAGT
CCATTGTCAAATATGTGATTTACAAATATTTTCTCTCAATCTGTAATTTA
GTTTTCATCCTCTTAACAGGGTCTTTTGGAGAGCAAATAATTTGATTTTC
ATAAGGTTCAAATTATTAATTTTTTCTTGTATAGTTCACACTTCTAGTGT
TAAGTCTAAAAACTGTGCCTTGTCATAGGTACCAAAGGTTTTCTCCAGTT
TTTTTTCTAGAAGTTTAGAGTTTCATGTTTTACATTGGAGTCCATGATCC
ATTGTTAATTAATTTTTGTATATAGGTAGATGTTTAGGTTTAGGGTTTTT
TTAAAAAAAAATTACATATGTTTAATTGCTCCAGTTCCCTTTCATTGAAA
AGGGTATCCTTCCTCCATTGAATTGCCTTTGTCAGAAATTAATTGGACAT
ATTTGTGTGAGTCTATTTCTGGGCTCTTTATCATGTTACTTTTAAAAAAT
GCATCAGTTCCTCCACCAATACCTCATTGTCTTGATTATTGCAGTTATAT
AGTAAGCCTTAGCATTAGGAAAAGTGTTTTTCCTGCTTTATTCTTTNTCA
AAAAATTTTTGGATATTCTAGGGCCTTTACATATAAATTTTAAAATAACT
TTGTCTATGTCTAACCGAAAGCCTTATGAAGATTTTGATAAGAATTGCAT
TATGCCTATACATTAATTTAAAAAGAACTGATGTCTTTATTCAGTTGATT

FIG. 3 (3 of 52)

CTGCTAATCTATGAACATAGCATCTCT..CAAAGCATTTAGTCTTTCTT.
AATTTCTGTCATTAATTTTTTAAAATTTTCATCCTAAAGATTCTGTATAT
GTTTTGTTGAATTTATGCTTAAGCATTTCACTTTCTTGGTAACAATTATA
AATGATTTTGTGTTTTTTATTCCACTAGTTCATTTTCAGTGTGTAGAAAA
GCAATGAATTTTTGTGTGTTGATCTTTGTTCCTACATCTTGCAACATTAT
TGAACTCATTTATTAGTTCTAGGAGGTTTTTTCATTTTTCTTGTAGATAC
CTTGAGATTTTCTATATAGACAGTCATGTTGTCTGCAAACAGGCACAGTT
TTATTTCTTCCTTTTCAATCTATATGCCTTTTTTTTTTTTTTTGCCTTAT
TGCAGTGGCTAGAACTTCTAGCACTATGTCAAATAGCATTGGTGAAAGCA
GACATCCTTGTTCCTTGTCTTAGAGGAACATTTGGTCTTTAATCTTGGAT
TGCG
>Contig13
GCGCCTCCTTTTCTCTTCCAAAATTTCTCTTGTCTAGTTATTTGTCCAGG
GAAATTTGAAAGCTCACTTACTGTGCAAGTCAGCAGGAAACAACTGGGTC
TGTGCACAGCACCTAGCAAAGTTCTGCTCTAGGAATTACACTTTGGCCCT
GAGGTAGATTTCTACAAGAACCTTACCTTCTAAGCAGCACTGGGGTTCAT
CTTTTTCCCAGTCCTCAGAGCCCATTTTCACTCCTGAGTTCTCCCCCACA
AAGGACATTTTCAACGTTGAGTTTATTACTCAACAGAAAATGGAATGAAG
TCCAAGACCTAAGGAGATAGAAAGGGGACCAGTTATGGCATCTTCTCACC
CCAGGACACCTTGCTGCATGTCTCTAGTGCTGAACAGACCACTGGCCTTG
CTCTGTAGTTTGAAATGCTCGCTGCAACCAGAAAGGCACCAAGGGGCCAG
ACCATGCTCTCCTGTCTATCACGCCTTCAAAGCAGAATTTCCCAAACCTT
GAGTCACAGTGCTAACACACGGGGTGCCATAACATTTTTGTTGATTTTGG
CATTTTACAAAAATAAAATAAAAAAGTTAAAAATGCATTGCTCTATTCTT
GGGGCTGGCACACTATTGCCTTTGGCCAAATCCGGTCCCTGACTGTTTTT
TTAAATAAAGTTTTATTGAAACACAACCATGCTCTTGTGTACATATTGTC
TCTTGGCTGCTTCGAAGCTACAATA
>Contig14
GTGTTCGCTTTTTAACACTTACCTAAAATTACTCTGTAATCCATGGATCC
TTAATTTATTTAAAAAACTAATGTTAATGAGTAGCTTTATTTTCCTCCCA
TCTAATTTAAGGCCCACAGAACACCTTCACTTACCTCAATCCTCTCCCAA
CTTACATGCTTTTAATGTCATATATGTTAATACCGTATACTTTTAAAACT
TTCTAAAATAGCATTATTTTATAGCATGAGTGTTCATTTACATTTTTGCA
TATATTTAGAATTTTCTTTGCTCTTCGTTTCTTCTTCTATTTATGACTCC
CCTCTGGGATCATTTTCCTTCTACTTGAAGTACATAGTTTAGAACTGCAC
TATTCAATACAGTAGCCACTAGCCATGTGTAGCTATTGAAGTTTAAACTA
AGTAAAATTGAGTAATATTAAAAACTCAGTTCCTTCATCTCACTAGCCAC
ATTTCAAGTGCTCAGCAGCCACGTGCGACTAATGACTACTGTACATCAAA
CATATAGAACATTTCCATCATGGCAAAGAGCTCTATTGATAGTGTTCATC
CAGAGTTTCTGTTCCAGGACCAAACTGAGGGTTGGGCTGCTATTTCTCAT
GGCCCAATAACAAGATGCAGATGAGCTGGGGAGGAAGAGAGTTTTTATTT
CTGCNACCATTTACCGGGAGAAGGCCTGGAAATCATCACCAGGCCAACTC
AAAATTATTACGTTTTCCAGAGCTTATATACCTTCTAAGCTATATGTCTA
CGTGTAAGTGTGCATTCACCTGAAGACGTTAGTGATTAACTTCTTTTAAT
CTGTAACTAAGGTCTGAGTCCGGAAGATCTTCCCCTGGAGCCTCAGTAAA
TTTACTTAATCTAAATGGGTCCAGGTGCTGGGGTAATTACCCTTATCTTG
TCCCCTGCTAAATCATGGAGGTTTGGGGATTCCTTTAGAGCACCAATAAA
CTTGTTTGTGGAGGCCTGGGGGTTTCTTCTGACCCACAATAAAACTTGTT
TAATCCTAAATGGGTCCTGTTAAGAATTCCTTCTTTATTTTGTCATATTT
TAAGGCCCAGAAAAGGCCTGGGCAAAACTCTTGATGGGCTTTTGTTACAT
TCCAGCCTTTGTATAAGAACACTGGTTTTTAATATTTAACTTAACCATTT
AGTCAGTACTGAAACAGTTGTTATAGAGATCTGCATTAGTGAGACCTGGC
CTGCCACATTTCCTTTTCTGAAGATCTTATGGTAGTGATCACCTTTGTGA
AAGGAAAATAAATCTTGGGACCTCAAAATCACTAAGCCAAAGAAAAAAGT
CAAGCTGGGAAGAATCTGACACTTAAATCCAACACTGCTAACTCATTCAT
CTCACTCATTCATTCATTTTATTTTCTTTTTTCTTTCTTTTTTTTTTTTT
TTTTTTGAAACGAAGTCTTGCTCTGTCACCCAAGCTGGAGTGCAGTGGAT
CTCAGGTCACTGCAACCTCCACCTCCCGGGTTCAAGCGATTCTCCTACCT
CAGACTCCTGAGTAGCTGGAATTACAGGCACCTGCCACCACGCCTGGCTA
ATTTTTATATTTTTAGTAGAGACGGGGTTTCACCATGTTCATCAGGCTGG

FIG. 3 (4 of 52)

TCTCGAACTCCTGACCTCGTGATCCGC...CCCCCTCGGCCTTGTTTGCT
GAGGTACTGTCTAAATGCTGGAACTGAAAATGGCAAGCAAGACATCCCTA
CCCTTGAGGAAACTGTAATCTAGTCGGAAATACAGATGTCAACCAAGTCT
CACACAAGAANATTGTACAAAACCCCTAGGA
>Contig15
GGAAAAACCTATCACCGCCTCCTATGGAACTTAAAACAAAAAGAAAAGTA
ACAAAGGAAATGAATATTTCATTCTGGAAGAACATTGAAAAAGAACAGGA
AGAAGAGAAAGCACAACTCGAACTGTCCACTAGAATTGACAACACTCTGA
CAGAATGTCTGAACCTCATCGAAGGGGTAAGTGAAAAAAATAAGCTCCTC
CAGCTTTGGCCCAAAGTCTTATAATTTTTAAACATATTCCTAAATATAAT
ATAGGAGAGATAGCCTTCATCTAAGTAGAAATTTAGCTACTCTTGTAAAT
ACAGAGTAATAATAATAATGACATGCCCATAAACAGTGTCTTTTGTGTAT
CTGTGCTTTTATAAGCACTTAGCTAAGATTATCTCACATAATTATCATAA
CCACTGTTACTATGACCACTTTACAAACAAAACTGAGGCACAAAGAAGTT
GGAAAACTAATCCAAACAAACTGGCTCCAAAAGGAACTTTGCTTTCTTTG
GGTATCAAGTTCTGAAGAGTACACATTTAACATTGAAACTGAGGTCAGAA
GGCAAGTTTCTATGTAAAGTTGGAGTATTCTGAATACTCTGGGTAGCTAC
AAATAGTATTTAAATTTTATCTTGGATTCTGCAGATAAGGATAAAATAGA
TGGTAGGCAAAGAGTATGATCCTTAGGAGAAATTTTTCCTGAAGGAAAAA
TATATTAATAAAAAATGATGGAATAAACTTCTAAGATCCTTGCCTAGAGC
AAAACTCATTCAGTCCTTTGGCTGGTAATGTTGAACATCAACAAAAAAAA
GGAAAAGTTCAGTTTAAGTCTACTCCAGGCAACATTTTCACAACATCCAG
TTAAATATTAACTATTTCTCTTTGTGGAATTGAACTAGAGTTCTTTTTCT
TATCCTCTTTTTTGGTTGTTGTATTATTTAAAAATGAGTACCTTTTTATT
ATTGAAATCATTTCAAGTAATGCAGATAAATGATCAGCCCTCTCCCTGTA
CAAACATACATACTTAGGCATCCCAAACTTCTCTCTGGAGGTGACCACCA
TTGCCAGTCATTCATTCTGTTTTCATGCATGTCCATACAGTATAGGTATG
TCGAGAAATGAAGTATTATATATTTTTGTGAGTTGCAATTCTTTTATTCACA
TTTTTGTGTACTTTGGTTGTCTTTTCTTGTGTTTTCCTAGTACCAATGTT
ATGCTGACTTAGGCAGATGAGTTGAGTATTTTCCTTTTTGCCCTATAAAC
TGAAAATAGTTTGTATGACATGAGAATTATTTTTATTTTTTGAAGGTTTG
ATAAAAACTTGCCCATAAAAATCGTCTGGACCGGTTTCTTGAGGATGCCT
GTGTTAGAGCC
>Contig16
CGCTTTAACCTGGGCTACCAATGGTTCGTCAAGTTCTAGATTCTCTATTA
ATACCTTTTTCTTGTGTCTTTCTCTGGTCTGTTTTCAGCCCCGAGTCTCT
TAGATCTGTCCTCTAATATTCCTATTGACTTTACTTCATTTTCTAAGTCT
TTATCCTTTTGCTTTACTTTCCGAGAGACCTGCTTAACCTTATCTCCCAA
CTCTTTTATTGAATTTCATTTCTTTTACTATATATTTTTTACTTTGAATA
CACCTCTCTCTTCCTCACATTTTCCCCCATAGTATTTTGTCTTCAATTGA
CAGTTCTACTATCTTATTACTCTGGAGATATTAATAATAGTTTTTAAATT
TTTATTTATTTTTATTTTCAAAACAGTGTCTTACTCTGTCACTCACGCTG
GAGTGCAGTGGTGTGATCATGGATCACTGCAGCCTTGATCTCTGAGCTCA
AGCTATCCTCCTGCTTCAGCCTCCCAAGTAGCTGGAACCACAGGCATGTG
TCACCATACCCAGCTAATTTTTTTGTTTTTGAGGTGGAGTCTCACTCTGT
AGCCCGGTCTGGAGTGCAGTGGTGCAATCTGGGCTCACAGCAACCTCTGC
CTCCTGGGTCCTGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGC
TGGGATTACAGAAACACACTACCATGCCCAGCTAATTTTTGTATTTTTGT
AGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCTGACCT
TGTGATCTGCCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAG
CCACTGCACCCGGCCACTAATTTTTAAATTGTTAATAAAGACGAGGTCTT
GCTATGTTGCCCAGTATGGTCTTGAACTCCTGGGCTTAAGTAATCCTCCT
GCCTCAGCCTCCCAAAGTGTTGGGATTACAGGTGTGAGCCACTGAATCTG
ACATTTTTTAAAAGTTTTCTTCTCTTTACCAAGTCTTTTTTCCCCTTTCT
GCTTTTTTGGGTTGTTTTATTTTGATCTCTATCTTGCTAGAAACTTTCTG
CAGACGTTTAGTAATACTAGATTTTTGAGAGTGGGCAACTGGAAAGCTGA
TTGGAAACTCTGAATACATGGGTGAGGCTTGTTGGCTGTGAGTGTCATTG
CTTGATGTCCTGGCAAGGCCAATGGGTTTGGACCCCTACTATTAGTATA
GGCCTGATTCCCTGGGAAAGGCTCTTTTGATCTCCTGCCTGGAGGATAAA
GGCCTGGCTACCAGCCTTCTGTGTGTAATGTGAGGGAGAAGGGCTGGAGT

FIG. 3 (5 of 52)

ATTCAACATCATGCTGAA.CCTTTCAA..ATCATCTTGTTTTTAGTAATC
TCCTACCTTAACTCTCTGTCTTCTGCTAGTATGGGAAAGATGACCTGAAA
ATCTAACCATTTATTTTTCCCCCATTAATATCATTTTATGATTATTCAGA
AGTTAAATAATTGTCATGCTGTCCTCCAAAAAGACTGAATCAACTAGCAA
CAAATAAGAATTTTCTCACAGCTCTGCCAGCATTTTAAAAGAATAGCTTT
ATTGAGCCCAGGAGGTCAAGGCTGCAGTGAGCTGTGATTACACCACTCTA
CCCCAGCCTGGGTGACAGAGCAAAACCCTGTCTCAAAAAAGAAATTTAAG
GAACAGCTTTATTGTTGTAAAATAGACATACAATAAACAGAGCACATATT
TAAATTGTGCAACTTATACTTTGATATAACCCTGTGAAAACATCACCACA
ATCAAGATAGTGAATATATTTATCACCTCCTGATACAGTTTAGCTCTGTG
TCCCCACCTAAGTCTCATGTTGAATTGTAATCCCCAATGCTGGGGGAGGG
GCTTTGTGGGAGGTGATTGAATTGTGGGGGTGCACTTCCCCCTTGCTGTT
CTTGAGATAGTGAATGAGCTCTCATGAGCTCCCCTTCACTCACTCTCTTT
CCTGCTGCCATGTGAGGATGTGCTTGCCTCTTCTTTGCCCTTCTGCCATG
ATGTGTTTCCTGAGTCCTCCCTAACCATGCCTCCTGTACAGCTTGCAGAA
CTGTGAGTCAGTTAAATCTCTTTTCTTCATAAATTACCCAGTCTCAGGTG
GCTCTTTATAGCAGTGTGAAAAGGAACTAATATACCTCCTAAGTTACCTC
AAGCTTGTTTTAATTCCTTCTCCTCCCTTCCTTCATTGCCAAGCAAACA
ACCACCTGTTTCTGTCACTATAGATTAGTTTACATTTTGTGGGTTTTTT
TTTTTTTGAGACAAGGTCTGACTCTGTTGCACAGGAGCAGAGCAGCGTA
TC

>Contig17

CGCGTTATAGGAGATGCGAACTTAAGAAATGATGATAAGGAGACTTTATT
AAATATAATTTTGAATTATTTTGCCATTACAGAAATTCTAATTATTTAAA
ATTCTATTCATAATTTTTAATCACTGTACTTCCCAAGCTTAGCTTAGAAT
CCTTCTGTGCTGAGGATTAATTTTAATTTGTCTTTTATAGGCCTTATCTA
AAATCCAAGAATAATTGCCAGAATCAACCACCTTCTAAATCTGTAAGTAG
AAATTAGTCTTTTTAAAAATATGCATTCATAAGTATGATTAGTAATAAAA
ATAATAAAGATGTTAGCAACCTAAAGAACATGTATTTGAAAGGTATTTCT
TACAGATATAAAAACAGTTTGGTTTAATAAGAGACAATCATTTTTTGAAA
AGTATGACATTTTTTGAAAAGTAGTTTAGTTTTATTAACCAAGAAAAGCC
TCAAGTGAACTTTAGTCCTCTTGATAGCTAACATTTATTGAATGCTTACT
GTGTGCCTGATACTTTTCTGACTTGCATTACCTCACTGAGTCCTCACAAT
CTTATGAGGCTACTATTAGTAGCCCCACTTTACAGATGAGCAAACTAAGT
CACAGAAAGGTTAAATAGGTCGTATAGCTATTAAGTGACAAAGCTGAGAG
CCTGTGATCTTAACCACTTTGGTATGCTGCCATGAAGTTAAATAGCTCAG
TAGTCATTAAAAGAGAACATTTGCATTGAACCTTCCAAGCCACTTAACAA
GTATATGCTTCCTAATCAATTTAATTTAGCTACATTAGATAGAATGGTAA
AGGATCCTTAACTTAAAGTTTAAATGGAAGAAATTAGCCCTCTGAAAGAG
GCACAGATTATTCATCTGCAATAAAAATCTCACCTTTAGTTTTTTAAAAC
ATAGTTTTTATCTGTGTTCTGAAATGTAACTAAAACAGTGCTTCCTGAAG
TGAAAAATTCTCACTGGTGAGAATTTTAATAAGTTTTAATGATTCACCAA
ATCACTTCAGTCATATTTCAGTCATATGCATATGCATATATAGACATATA
AGTTTTTATCTGTGTTCTGAAATGTAACTAAAATAGTGCTTCCTGAAGTG
AAAAATTCTCACTGGTGAGAATTTTAATAAGTTTTAATGATTCACCAAAT
CACTTCAGTCATATTTCAGTCATATGCATATGCATATGTAGACATATATA
TGTTGTATGTATACATGACATCATTAGACACTGTGAAGGATAGCAAAATG
TATATAAGGCAAAATTTATGAACAATGGTTTAACGTTTGGGAAGCACTGG
GTTACACTTTTACTTTATGCAGATTGAACCAGTATAGTATGCAAGTCTTA
AGGAAAAATCTACTGGAAAGGGCCCTCATTCAGACTTCCCAGAGGCTTCT
CTGGAAGTTGACAATACTGACTTCAGTACATCAGCTCGTAAATGAGGATG
ATACCTACCTTATCTGCTTTACACAGTTGTAAAAGTAAAAAGTGAACTCA
GGAAGGGAATTACAGAATTTAGGAGAAACTAAAAGCACGATGTAAATAAT
AGTCATCATTACAGTTATATAATGCTTGACAATTTATATAACACTTTCGA
TACATGACAACAATAACTAACACCCAGACATGTTTATATACATTACCTCA
CTCAGAACAACCATGTGAGGAAGTTGGCCATATGCTTTAATGTCCAAACC
AGGACACTTTTGAGAGTAAAAAGCAGTACTCTTTGACCAACAGGCATAAA
TCAAAACTATCTTGTGAAAACCGGGATATATGGCATCCTTCCTAGATAAT
AGATACTTTTACTATTATTAATTTTGCTGTGAATCTAAACCTGCTCTAAA
AAAGTTAATTTTAAAAAGTAATGAAGTACTGATACATGCTACAACATGGG

FIG. 3 (6 of 52)

TAAATCTTGAAAACGTTATGCTAAGTG..AGAAGCCAGACAGAAAAGGC.
ACATATTACATGATTCCATTTATATGACACATCTAAAATAGGCACATCTA
TAGACATACAGAGACAGAAAGTAGACTAGCGGTTGCCAAGAACTGCAGGG
AGCAGAAGATGGGGAGTGACTGCCAATANGAAAACGCATTACGT
>Contig18
TGAATCGCAATGATATGTGCCACTTTGCACTCTCTGTGACATATATAATT
ATTTTTAATGCATTCATTTTTTTCTCAGAGTGCATTCGTTTGAAAACATA
GACGGGAAATACTGGTAGTCTTCCTTGTCAGTTAGAAACACCCAAACAAT
GAAAAATGAAAAAGTTGCACAAATAGTCTCTAAAAACAATGAAACTATTG
CCTGAGGAATTGAAGTTTAAAAAGAAGCACATAAGCAACAACAAGGATAA
TCCTAGAAAACCAGTTCTGCTGACTGGGTGATTTCACTTCTCTTTGCTTC
CTCATCTGGATTGGCATATTCCTAATATCCCCTCCAGAACTATTTTCCCT
GTTTGTACTAAACTGTGTATATCATCTGTGTTTGTACATAGACATTAATC
TGCACTTGTGATCATGGTTTTAGAAATCATCAAGCCTAGGTCAGCACCTT
TTAGCTTCCTGAGCAATGTGAAATACAACTTTATGAGGATCATCAAATAC
GAATTCATCCTGAATGACGCCCTCAATCAAAGTATAATTCGAGCCAATGA
TCAGTACCTCACGGCTGCTGCATTACATAATCTGGATGAAGCAGGTACAT
TAAAATGGCACCAGACATTTCTGTCATCCTCCCCTCCTTTCATTTACTTA
TTTATTTATTTCAATCTTTCTGCTTGCAAAAAACATACCTCTTCAGAGTT
CTGGGTTGCACAATTCTTCCAGAATAGCTTGAAACACAGCACCCCCATAA
AAATCCCAAGCCAGGGCAGAAGGTTCAACTAAATCTGGAAGTTCCACAAG
AGAGAAGTTTCCTATCTTTGAGAGTAAAGGGTTGTGCACAAAGCTAGCTG
ATGTACTACCTCTTTGGTTCTTTCAGACATTCTTACCCTCAATTTTAAAA
CTGAGGAAACTGTCAGACATATTAAATGATTTACTCAGATTTACCCAGAA
GCCAATGAAGAACAATCACTCTCCTTTAAAAAGTCTGTTGATCAAACTCA
CAAGTAACACCAAACCAGGAAGATCTTTATTATCTCTGATAACATATTTG
TGAGGCAAAACCTCCAATAAGCTACAAATATGGCTTAAAGGATGAAGTTT
AGTGTCCAAAAACTTTTATCACACACATCCAATTTTCATGGCGGACATGT
TTTAGTTTCAACAGTATACATATTTTCAAAGGTCCAGAGAGGCAATTTTG
CAATAAACAAGCAAGACTTTTTCTGATTGGATGCACTTCAGCTAACATGC
TTTCAACTCTACATTTACAAATTATTTTGTGTTCTATTTTTCTACTTAAT
ATTATTTCTGCAATTTTCCCAATATTGACATCGTGTATGTATTTGCCATT
TTTAATATCACTAGACAATTCAATCAGGTTGCTACGTTGGTCCCTTGGGT
TTACTCTAAATAGCTTGATTGCAAATATCTTTGTATATATTATTGTTTTT
TCTCCTATCTTGTAATTTCTTTGAGCACATCCCAAAGAGGAATGCCTAGA
TCAATGGGCACAAATAATTTGACAGCTCTTATTAAACATTATTCTGTAAG
TAAAAACTGAACTACTTTTCAGTATCACTAGCAACATATGAGTGTATCAG
CTTCCTAAACCCCTCCATGTTAGGTCATTATGAACTTATGATCTAACAAA
TTACAGGGTCTTATCCCACTAATGAAATTATAAGAGATTCAACACTTATT
CAGCCCCGAAGGATTCATTCAACGTAGAAAATTCTAAGAACATTAACCAA
GTATTTACCTGCCTAGTGAGTGTGGAAGACATTGTGAAGGACACAAAGAT
GTATAGAATTCCATTCCTGACTTCCAGGTATTTACACCATAGGTGGGGAC
CTAACTACACACACACACACACACACACACACACACACACACACACACAC
CATGCACACACACAATCTACATCAACACTTGATTTTATACAAATACAATGAA
TTTACTTTCTTTTTGGTTCTTCTCTTCACCAGTGAAATTTGACATGGGTG
CTTATAAGTCATCAAAGGATGATGCTAAAATTACCGTGATTCTAAGAATC
TCAAAAACTCAATTGTTTGTGACTGCGCAAGAAGAAAACCACCCATGCTG
CTGAAAGTCAGTTGTCCTTTGTCTCCAACTTTACTTCCTTTACCTCTCAT
ATGTTTGTGAATAAGCCCAATAAGCAGACNCCTCCTACAAAGTGAACCTG
GTCTCTTTCCTCCTAACAGGG
>Contig19
GTCTTGTAACACAGGTAAGACGAGTTCAAGTTTTATTTCTTGNTTTTAGA
ACGGTAGTGAGCGGTTTTCAGCNTGAGACCACACCTAAGGTAAGTAGCTG
AATTGGGGTTTTGTCTTGGCTAAAGTTTAACAACCAGCTGGTCTTAATTT
CTCCTTACCATTAGAGCACTCAGTAATCATATAAGTTGTGTGATCATTCA
TTTTGCTTAACTGTTTGTTTCTGTTTTTATTGCTGTTTCAGTCTTTTTCC
CATTGGGTTTGACCTACTCTATCTGACTTGATCAAATCCAAAGGAAATTT
CCAAATTATGGGGAATGAGGCCTCTGAAGTGGCTAAATTCCCACCCTCCC
ACACACACAAACGTGGTATGGTGGGGGAAAAAACGGCCAGCAAAAGAAAA
AAAAAAAGGAAAAGATGTTTCATTTTGACCACCAAACGGGCTTTATTTAC

FIG. 3 (7 of 52)

ATAACAAGGCCACCTTT..GCTAGCCA .CCATACTGAAAGAGCAATGC..
TGTTGCCCCATGCTGTGGGTTCCATAGCTAACGTTCTGCCTTTTTTCCTA
CCACGACAGCCTGGGTTTGGTTCCTAAATCAAGCCTTTTCTGGTTTGATA
CTTGGTAATGCTGAAATAGCAGCAATTTGTCCTAGCTGAAATATCGTAAT
AAGATTTTAAAAGATTTATTTTAAAGGACCTCAATAGTTAAAAGTCAGCT
TAATTAAAAGCTAACATCCAAGATGTGTGCATGTGTATGTATGCGTCTTT
GTATTTAAATAGCCCTCATGTTTTTTTTTTCTTTCCTAGGAACTTGCCTT
TTTTTGAGCAAAAGTTTTTTTCTTCTCTGTTGACTGGATTCTGTTTTCTT
CATTTACTTCTGCTGTCTCTCCTTTCTCTTGCACCGTCTGCTGCATGAGA
GCCCTAAAATAGTTTATAATAGCCTGGGGTTCCTTAAAGAAAATGGAGAA
GGTGCCAGGCTCCCTTTTAGGGAGAAACTTCTATTTTTCCTTATGGAATC
CCTAGAGTGTAAACAGACAAGTTCATTTCAGCTCTTAAACTGCTTGCGTT
TGTGTTGTGTTACCTGATTTTTTTGACTATTATATTTTTGACTAGCTATT
GCAACAGAAGCTACTCTTGGGTTTTCAAGGAAGATTGTAGTTTAGACATG
TAGAAATGTCTTTTAAAAAAAAAACAAACTTTTTTTTAAGTGCACTGTAA
AAGCATCATATGGTCTAGCCTCCTAATAATTTTCCCTTTTTGGAGACCAG
GATTCAGGGTGGGCTCTGCCCAGAGCTCAGAGATCCAGTTAAAAGAGAGG
TAGTCTCGGCCGGGCGTAGAGGCCCAGCCTGTAATCCCAGCACTTTGGGA
GGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCTGGCCA
ACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTG
GTGGCAGGTGCCTGTAGTCCCAGCCACTCGGGAGACTGAGGAAAGAGGAG
AATCGTTTGAACCCGGGAGGCGGAGCTTGCAGTGAGACGAGATGGCGCCA
CTGCACTCCAGCCTGGCGACAGTGAGACTCCGTCTCAAAAAAAAAAAGAT
AGGTAGACTCGATGTTGTCGTACCCGAGCAAGTTAGAGCAACGCCACACT
TTGAGACGAATTTAAGAGTCCTTTATCAGCCGGCGACCAAGAGACGGCTA
ACGCTCGAAATTCTCTCGGCCCCTTGGAAGGGGCTTGATTTTCCTTTATG
CTTTGGTTTAGGAAGGGGAGGGGAGCTCAGTTGCAACAATTCTACAGGAG
TAAAAACATGCAAAGAAATTAAAAAGACAAGTGGTTACAGGGAAACAAAC
AGTTCCAGGTGCAGGGGCTCTAAATCTATCATAAGATGTTAGGTATGGGG
GCTCTGCCGGACACAAACTCAAGGCTTTATGCTGTTATCTCTTGAGCGAA
ATCCTGGGAACTTCGTACATTGCTTGCTTCAGTACCTTATCAGTTAATCG
GACTCTTTGATATGTTGGGAGTCAGCGTACACAAGTTAACTCCTTGAGGA
AGGGGGTGGGTAAGGAGTCCTTGATGTCTGGTAAATGAAGGAGCGAAATC
GAGTTCCTCTGGCTTTCTCAGCTAAGGGAGAGCTTATTCATGTGGAAACA
AGGCTAAGTGATTAAGGGAGAAAGGGAGAGTCTGAAAACAAGGTTAGGTA
TTACAATGTCAATAAAATTGGTCTCCTTATACAGTCCTATGGTAGATTTC
TTTCCATCTTTAATCTCCCTCTAGCACCACCAGACTTTTTCTCTCTGTAC
CTTGAGATGTAAATTTTGCTATCTGAATTTTCGTCTAAGAGTTGTTTCCT
TTAATATGCAAATTTAGGGTTATTTAGCTGACAACTGCCAAAGTAGTGAA
ACAAGTTATCAAGAACTTGAACGTCTAAGGTAGGAAAAAAAAAAGTCTTT
ATGAATCTATAAGATGTACTTCTATTGGCATGCCTAATACGTCTATGTAT
TTACGTGTTGTGTACACAGTTTTTCACTACTGAAAATATATAGAGGAGTT
CTAATTAATTGACTTAAGACAATAAAAGCGCTTGAATCAAATACCTTATC
AGGAAAAAGGAAAAGACAAGTCAAATGCTTGTTCAAGTCTATATAACTTA
AGTAAAATCTTTAATAAATAAGCTAGCTTTAACATTATTTGAAATGTCTT
AAGAATTGCCAGCAGGTTCTGGGTTACAGAACTAGTGGGGGTGCAGTGGG
GTGAGGGTTGGTGGGGTGGGNGGTNNNACNNNNNCNCCCCCCCCCCCCCC
CCCCCCCCCCCCTCCCCCCCCGCCCCGNGCGGCCGCGCCCCCCCCCGC
CCCCCCGGCCCGCCCCCCGCGGCCCCCACCCCCCCCCCCCCCCCCCGC
GCCCCGCCCCCCCCCCCGCGCCCCCACCCCCCCGCCCCCCCGCCCCCCC
CCCCCCCCCCACCCCCCACACCCGGCCCACACGCACCCCCCACCCCGAC
GCCCCCGCCCCCCCCCCCCGCAGCCGACGCCCCCCCCCCGCCCGCCCCG
CCCCGCACCCCCGACCCCCCCCGCCGCCCCGCCCCCGCCCCCCCCCCCCG
GCCCCCCCCCCGCCGGCGCGGCGCCCCACCCCCCCCCCCCAGCCCCGACC
GCGCGCCCCCCCCACCCCCCCCCCAGCCCCCGCCCCCCGCCCCGACCC
>Contig20
GGCAGTACGCTATAATTCCCTCTTCACCTTACCTCATCTGTTCTCTGATG
GATGTACTTTTTTTTTTAGTTTCTAAATTCCCTTTTCCTTTGCTCTGGAG
ATGGGTGATTGATGTAGTCTGGGTATTTGTTCCCTCCAAATCTCATGTTG
AAATGTAATCCCCAGTGTTGGAGGTAGGGCCTGGTGGGAGGTGTTTGGAT

FIG. 3 (8 of 52)

10/118

CATGGGGGCAGATCCC. ATGAATAGC GGTACTGTCCTCTCAATAG
AATGAGTTCTCCTGAGATATGGTTGTTTAAAAGTGTGTGGCACTCCCCCA
TTGCTCTCTTGTTACTGCTTTCGACATGTGACATCCCTGCTCCCCTTCGC
TCTCTGCCATGATTGAAAGTTTCCTAAGGCTTCGCCAAAAGCTGAGCAGA
TGTGGGTGCCATGCTTGTACAGCCTGCAGAACTGTGAGCCAAAATAAACT
TCATTTCCATATAAATTACCCAGCCTCAGATATTTCTTTATAGCAACATA
AGAGTGGCTTAATACAGGCTGGGCATGGTGGCTCACGCCTGTAATCCCAG
CACTGTGGGAGGCTGAGGGGGGTGGAACATGAGGTCAGGAGATTGAGACC
ACCGGCTAACACGGTGAAACTCCATCTCTACTAAAAATACAAAAAATTAG
TCGGGCGTGGTGGTGGGCGCCTGTAGTCCCAGCTACTCTGGAGGCTGAGG
CAGGAGAATGGCATGAACCCGGGAAGCGGAGCTTGCAGTGAGCCGAGATT
GCACCACTGCACTCCAGCCTGGGCGACAAGAGTGAAACTCCATTTAAAAA
GAAAAAACAAAATTTCAAACAGAACAAAATGAAAAAAATACCAAGTGAAA
GGCCCCTATAAAAACCCCTCTGGGGCCCATCCTCCCACCCCCTCAAGTGA
AACCACATTTAACAATTTGGTGCATATCTTTCCAAACCTTTTGTTGTACA
CATATAAAAAACATACATGCTTTGATTTGGCTCAGACTGTACATAGTGTT
TTCCCTCTTGCATTTTACACTTAATATATCTTTGACATCTTTCTATGTCA
GTGCATGTTGGCTCGATGATATTCTATCATTAAATACCCTTCCAAAAATG
GTAAAATCATTTTAAAAAATCATTCACACAAGTACATATTTACAATTTTA
AAAGAAAACAGAATCCCAAAACACAACGACAAACCTCTAAAAATAATCTC
TATCTTTCCACCAGCATGGAACAGTTCATTCCTTTTTCACATAAAACGAA
TTATGTGATTGGAAAGATTAACTCTAATCTACACATTTATATACAGAATG
TTCTATTTGTTAAGCCTATCTGAAAATAAAAAATTCAGATGATTAATTCA
CTTACACTTAGAAATTAAGTCAATATACTATGAATACACATTGTGATCAG
TTATAATATGATGCTTCTTAGTCTAGGGTTTCAATTAAATAACAGTAAAA
AAAATTGGATAAATAAGACAGCTAATAACTGAAAAATCCAGAAATTCAAA
GATTATATTGCCAACTAAAACACTGCCATTTACATTTTTTTTCCTACTT
GGTAGCAAATGCTAATGGAATTCAATCCTGATTACTTAAAGTCAGTTCAC
ATCACACATTCAATCAGGATAATACGAACATAATATGCCTACTATAGCGT
TAGATTAAGACATAAAATTTTTTTGCTTGAAAGTAATGACTGCGTACCAC
TTGAGACATTTGTCAACCACTTCAGCACATTGTTTACGAGTGACTGGATG
TCCACAAGGAATAAAAACGACAGCAATATTTCTATCCATACAGATTTTGC
AAAGCTTCTCCTCTTGCAGGTGTCTTAGCTGCTCTTCAGTACTAATCTCT
TTCTGCAATGAAGTCTGACTTGATTCGTCTTGTGTACTGTCTTTCTGAGC
CTTCACTGGATCTGCAATCAGAACCTCAAGTGATTTACAGTTGCTCCCAG
ATGTCTGAATTTTTTCCTCCATTATTTTCTTAATGTCTTTGAAACTGAAC
CCCATTCATATAGCTTCTTGTACCATAGGATTATGGAAGATGGTATCAAT
TTTTCTAGTTAGTGATGGCGTTTTTTCAGCAGTTCTTACCAGACACTCCT
CAAGTGAATGGGATAAATGAATATTGTTTATATATTTTCGTGTCTTCTGT
TCTAACAGATATTTACACCCTGGATGCCATTAACATGTTGTCCCAAGGGT
CTTNCTGGGCT
>Contig21
CTTTCTCCCTTTTTACCCCCATTTTCGTAGGGATTTGGTTAAAACCCATG
TAAAAAATCCAAACACCGGCGGGGAACGGGGGTTCAAGCTCGTATCCCCA
CCACTTTGGGAACCCAAGGTGGCAGGATTGTCGGAAGCCAGGCATTTGAG
CCCACCCTTGGGAAAAAAAAGAGAACCCCCCATTTTTTTTGAACAAAAACC
CCAACCCTCCCAGGAAAGAAATAAGTATGGCTGGGTTGAAGTCACCAAAG
ATGGCCGACTGGCTGGTCAAGTAACTTTACCTGATGGTTCGTAGAATATT
TACCTTCACCCAGGTGGGAGAATTGCTTGAGCCAACCCTCAGTGTGGATT
CAGGAACTTGATTTAATTGGTATCGTGATTGTGGATTAGATTCTCAGGGA
TGCATTCACTAAGTAAAAGTGATAATAGCTACTTTTAAGTAAAATAATGA
ATGAATCAAACACTCTAAATCCATGGTGCTATGCTAAGCTCTTTCTGTAT
TTTATCTCATTTGATATTACAAATATTTGATGTGTTAATAGTAATGACTA
TCTCCATTTTTACAAGTAAGGAAACTGACATTGAGAGATTAAAAGACTAG
CACAAATCACAAAGTAAATGAGATTTGAATCCGGTCTTGATTCCAAACTC
TACAGTATTCTAAATTCAAGGAGACTAAATTATAAGATGGAGAGCCAATT
TTACTTTATAACAGGGTTAGAATGGCAGAAGAGACCTGACATTCACACCT
CTAGCCAGTGCATCATCTTCCTGTAGGCAAATATGCAGGAAATCTATAAT
AAGAACGTCCTTTGGTGAAGGCCAGGTGCAGGGGCTTACACTTGTAATTC
CAGCACTTTGGGAGGTCAAGGTGGGAGGGTCGCTTGATGACAGGAGTTTG

FIG. 3 (9 of 52)

AGAACAGCCTGGGCAACATAGTGAGACCCTGTCTCTACAAACAAAAACAA
ACACAAAACAACTTCAAGAAAACTCCTTTGGTATGGATCAGAACAAGATG
AATTATCTATCTGATCCAAATGCTTAATGACATTAAGCCACAGTCCACTC
ACTGCCACAATAGAGATATACCTGCCAATGCCACTCAGGTAATCCCATCA
AAAGTGGTAATGAGGTCTGCAGCATGACTTGTTCTTAGTGATCCCAGCCT
GAGACCTTGAGATTGCAGCATTTTATTCTACATATGCACAAAACATCTGT
TGAAAAAATCTTCTAAATTGATGCAATACATTCGTATCAAGAATACCTGTC
TGTAATCTCCATAAACCCTCTCCTTTCTGTTTTAAAAAATAGTAACAGCA
TTTCTCCTTACATGACAAAGAAATGACTTCACCATCTACGAAATAGTGAA
TAGGAGCTGTGTGGAAGGAAATTAGCTCTACTTCTTGGTGGAGATGAGAA
GGGAGTGTTCCTCTGAAAATCAAGGCTCTTGTCATGCTAGGAGCCAAAGT
CGTTTTTTAGAGTGTGGACAGTTGAGAAGATAAGACAGGGACCATCCACT
CATGTTTTTCTTATTCCATAGGCCTCTCTCAATTGGGCAAAGCACTCCAG
ACCTTTTGGAAGAGTGACACCAAAGGCAAGCACCTGCTTGGCAGGCCCCT
CAGCTTCTACGCAAGTATAAGTGAGTATATAAAATGGGGGTACTTGTGCT
GTTGAGTACCTTATTTCCAAATGAGGCCTGCCGGTGTCCCTGTGGCTGTG
AGAAGGCCTCTACTGGATAGGTGGAAGTTGTGTGTTCTCATCTTTTCTAA
CCCTGGATTGACTTGCCCAAAAGGAAGCCATTATTAACACTATAATAAAA
CCATCCTTAATCTGGGACTCTCTTCATGCAGTGGTTCTTAACCAGTGATA
AACATGAGAGTTACTTTTGGAGCTTAAAAAAATTAAGATGCTCAAGGTCT
ACCCAAACTGACTGAATCTCCAGAGGTGAGGCCCAGGGATGTATACTTTT
GAGCCAGACCTCAGTTTACCCTGCAGAGCTCATAAGGTTGCATAACACCC
TTTGTCAGCCACTCTGATGAAAAGAAAAATTGGTGAGGAATAAGTTTTAG
AGAAGAAGGAGCAAAGGTGTTCTTGGCCAGTGAGAGCCAATGACAGGGAA
ATGCAAACAATGTATCCACAAGAAAGGTAAATTACCCTATAGAGCATTTT
AGGATAAATGAACATCTCATGCCTAGGGTTGAGAGAGGGTACAAAAAAAA
AAAAAAAAAGACCACTCTGGATACACAACGCGATAAATGGAATAAAGAA
TTTTTCCTTGTAAATTAAAAAAATCCTTTGTTACTGAGGTATAATTTAA
TCTATTTTATGTATAGTTCAATGAGGTGTTATAGATAATAAATTTTTTTT
GTAAATTATTATATTGTCATATACTCATACATTCATTTTTAAAAGTCAGA
AATGTATATAACCATTAAACTTATAAATCATTCAGTCATTCAGAGATATA
GATACACGAGCATATTTTATATCCACCACAATAATTATTACCATCTCAAC
AATTCCATCACCCCTCAAATTTCAAGCGTAGGGGTTTTTAAATGTCAAAG
GAGTCTACTCAGTGGGAAGAAAGTTAAGGAAAAAACCTTTGGGGCTTTGG
GCTCCTTCCCCCTGGGGTTAAAAAGGCAGGAAATTGGGCTTACCCCCCCT
GAAATTGGGAACTGAAATTTTGGGAAGTTTAAAAAAAAAAAAA
>Contig22
TCAAGCAGCCTTCCTTCCTTGGCTTCCCAAATTGTTGGGATTACAGGCAT
GAGTCAGGATTCCTGGCTTAGTTTACATTTTCTAGAGTTTTGTATAAATG
GAAACATACAGAATGTATTTTTTTGCGGAGTGGGGGAGTGTTTCTATTTC
TTTCTTTCCATTTTCCCCCCCCCNCCCCCCCGAGACGGAGTCTCGCTCTG
TCTGTTGCCCAGGCTGGAGTGCAGTGGTGCGATCTCGGCTCACCGCAAGC
TCCACCTCCCGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCT
GGGATTACAGGCGCCCGCCACCACACCTGGCTAATTTTTTTTGTATTTTT
GGTAGAGACGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCCTGA
CCTCGTGATCTGCCCGCTTCGGCCTCCCTAAGTGCTGGGATTACAGGCGT
GAGCCACCGTGCCCGGCCCAAGTGTTTCTATTTCTTAACCAGCTTTCATG
CAATCTTTTTTATTTACCATCTCTGTGATCCCACTCCCAAAGGTACTA
GATGTCGATTGGTCCTTAGGATCAGCTACCATTTGCCCAACTGCTTTCCA
GCCTTCCAAAAATTTTTTTCTTTTTTTCTTAAAGATACTCCTGTGTGAGG
CTCAGAACTCTTGAATTGCTACTGCAAATATGAACTCGGTGATGTGAATG
CCAGGGAATTGCCTGATTGATCAAAGAAATGTATCCCCTTCTCCCTCACT
CTTGCTGTCTTCTCATTTGTTTTCCCCATCCTTGTGGATTCGTGAATTTA
AATATCCCTTTAATGTTATAATATTTTAATGGCGTTTGGCGAAAAGTACA
GAATTAGGTGCAAGAGTGCATAGCTGTTATTTTTTTTTGGCCTCTGAGA
CTGTTCATATATGCAAGTTATTTAACAGAAAGTTCTGCAGTGACCTGAGA
TGTCAGGGGGGTCTGATAGAGTACGTTTGAAGGCAGTTACTGGAAAAAAA
TAATGCCATTTCTGGTTTGTACTTCGGTAAGTTCAGATGACCCAATATAT
TGTTTACATGTGGCATTCAGTAAAAAGTAGCTTCCCCTCCCTTTCTTCT
TCCTTTTCTCCTTTCCTGCTTCTATAAAGCATCTGCTTTGGAAACTTCT

FIG. 3 (10 of 52)

TAGGAGGAGAGCTTGCCAGCCCGTGGC AATGGAGAGGTCTTGCAGAGA
AAAAGAGATGCTCCCACTCAATGCAGGATGGTGTGGAGGTAAATGGGGAT
ACGTCTGGCATCACTCAGGAATGGGCCTTCCTGGCAGGGAAAAAAAGGGA
GGGGAAAGAGGAAGGGAATTCNNANATNAATTGCTGAATACGGGGATTCC
ATGGCCTGGATCCAGGAAGAGAACTTTGGGAGGTGTGAACCTGGAAGGCA
TCANCTGATGAGGAGCAGCCTGAACTCCGGGGAGGACCTGTTTTTGGTGG
CCCGGAAAAAAATGCCTTCCACACACAGGGAGGCCACCCGGCTGATGGGC
TGGGGGTTGGACGGACAGCCCTAGGACAGGCTTGGGAAACCAGGCTCAGG
TAGGGCCTGCGAGGTTCTCGCTGCGTCTCTTTCCTTCCTGGTCTTAGAAA
ATAGAATCCAAGGCCTCTTGAGAGTGGAAGGTGGGTTGGGAGGAGGGCAG
ATGGGGCTTAGGCCCAGGACACCCGTAGAGCTACTGCCCAGCTGTCTCTC
AGGGACTCTGCTGAGGTCACTCCAAGGATCATTCTTAGCCTTGCTAGACA
GTACTGACAGAGGGAACCGTAGTATCGCACCCACTTCCTTCTCTTTCAAT
GAAAGTTTAAAGGTCACCATTTCCTCTGGCAAAGGAAGTTCCACAAATAT
TCCATTTCCGGTCTTAGAAACAGCAAGGTATCAAGCAATTGCAAACTTCC
TGTGCTGGGGAATTCCCAAGGAAGTAGGGGCAGAGTTCTGGTGGAGACAA
AGTGAATTCCGAGTGATTAGTCAGTAGCAGTAGCAGTAGCAGTAGCAGTA
GCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGCAGCAGAACC
AGAATTTCCCCGCACGTGTCTCAGGCTCTCATTTGCCAACTCAGTCTCTA
AGTATTTTTATTGGCAGGAAAAATAAAATAGCTATGAGTGAAATAATTCA
TTAGACCTGAGCCTCCATCAATTTTGTGTTTAAAGGCCTGACTCTCTTTA
CCTTTCCCTGGGATGGAAGATGCAAATGTTCCTGATCTCACTGTCAAAAA
AGAAGAACCAGTGGGTATATTGTATGCTTGAGTTCCAGCCATTAGTCACA
AGACATAGAGATGACTGCCATGTGTGTAGACTTTCTATAGACTGTGTGCT
AAACCCGACCTGCCACTTCCAAGGAGTAGATGAGGAATGTCCATGGTTCT
GGGGAGCCCTACCCCAATTTGGGGCAGACATTCCAAAGCTCATTTTCTGT
GGAGGGGGTTGATGGTTAAAGGAACGGCTGGGATTTACTCTTCTTTCTAG
GGCCAAGAAAATGACATGCTGCCTCCATGTTTAATCATCCTTCCCCCTGT
TAATAACTATGGCTTTAAGTCCCCGGTTAGGGCCTTCCTCCAAAATTGGG
GAAAAAAATTCCCCTCCCCCCCTAAAAATTTTTTTTTAAAAAAACCTTT
TTTTTTGGGGGTTGGGAAAAAAACCAAAAATTTTTTTTCCCCAGGGGTTT
TTTAATTTAAATTTCTCCCCAAAAAATTTGTTTTTTTTTTTCCGCGAAAAA
AAGACCCCCCCAAAAAAAAAAAGTTTTTTGGCGGAAAAAAAAATATTTTT
TTTGTGTTAAGAAATGGAGAAGAAGGGGGGTTTTTTTTTCTTCTCCCCC
CACCCGCCAAAGGAAAGGTTGTTCACAGATTGTTTTGTGTCTCCCGCCCA
T

>Contig23

ATGTGCCTGCGAAATCATCCTTCCAGAAATATTTGCCCCTTTCTTTTGTT
ATAGAGTGGCACTGCCCTATATGGTGACCACTTGCCACATGTGGCTGTTG
AACACTTGAAATTGGCTTGTCAGAATTGCAGTGTAAAGTGTAAAACACAT
ACCAAATTTCAAAGACATGGCACATAATAAAAAATGTAAAATATCTCATT
AACAATTTTTATATTGACTGTGTAAGTAACATTTTGAATATATTGGATTA
AATACATGGATGATGCCCCAACACCCACAGTCCCTTATCAAGTCTCTACT
TCACATTTTTGTACTTCTGACTTAGAAATAGCACTGGCGTCTAAGAGCCT
ATTAATGTCGTCAATAGGTTCTTGGGAACCACAATTTTAAACAAAATGAC
ATATAAGAAAACGAATAACATTGAACAAAATGACATTATTCGAGGACCTG
CTGCATGTTGTTTCACTTAAAGTCAGTGTCCAAGAAACTATCAGTGACAT
TTAGTGAGGAATTGCTGTCCTTCCTGTTTACAGGAACCTGGGCAAGTTAC
TTAATTCCTCTAAGCCCGGTTTATATCCCTGCAAAGAGAGAAGGATAATA
ATCACCAGTACTTAGTGATGTCGTAAGGAGAAAATAAAATAATAAATATG
AAATGGCTGACAGTGTCCTTGTCACACAGAAGATGTGTGATCCACAGTAG
CTGCTATTGTCTGCCTCACTTCACTAGTAATGGTCCAGGGAGGCCTTTAA
TGTGCATGGTGCAGTACATTCACATGTTGGACATGGGTGAAGGGAAAGAC
CAGGCTCATCTAAACACAATAGGATGCTTGTGGTGTTTTGAGGAGGAATC
AAGGACTAGTTATCCACAGCTGTAACATGCATGGATCAAAAGAGATAAGG
CACACAAAAGACTTTGTCAGTAGCAAAGCATTACAAAATGCAGAGACCAG
CTGTGGGTGGTGGTGAGTCAGACCCAGCTTCCCTCTGTGCCTGGCTGAGT
GGTTCTGGGCAAGTCACGCCATCTGTCTTGATGCCCTTCCCCATCTATAG
AGAGGGAGCAACTGAGGCCCCTTCCAATACTGAAGTCCTTTATTCTGCT
ACTTTAGAAATATCCACATTTTTGGTAAATTCAAATGATCCAATGATTCC

FIG. 3 (11 of 52)

ATTTCCTAATGTTCAAAACTAGCCCCACAACATCTAAATGAATCAAACA
AATAAAATATTTATTGTGTATGTTTTGATTGCTGAAACTTCTATTTTAGC
AACACACACACACACACACAGAACCCATAAGCCTTCATCTTTCCTTGGAT
AAACGAGCCTTCCTGTCTGGCCATTTAAGTCACGATTAAGTAAATGATTT
CCAACTCGCCTTTTGCAGCAGTTCAGATGGGTCTTTCCTGCGTGGCAGTG
GCCCTCCTGACTTATGATTTCCTGTGTGTCGGCCTGTTACCACTGCAGCT
TAACTGAGGAAACAAGAACAAAACAGCCTCTGACCCCAAGAGACTGTTGG
AGGCAAAGGCTTCAGTCCCAAGAACCTCACACGTGGGGAGCCCGAGAGCC
CAGCCCTGACCTTTTCTCCAGTAATAACATAAGAAACAACAGGCACTGGC
CTTATTTTGGATACAAAGAGTGGTGCTTTTCCTTAAATCTTCCTTTAGTC
AGGGGTACCCCTTCATGGACGCCCCAACATCCATGGTTCCTGCTTGAGTC
CCTGCTTCCATATTCCTGCACTTCTCACTTGAAATATCCCTGGAGTACGT
TAAGCAGCCAGGTTTGGAAGTTCTTGCTGTGCAGGCGGGTGTGTGCATGT
CCTCTCTCTCAACAGGACACAAGCTCCCCAAATCAGACGGTATGCCTCCA
CGCCCCTTCCCAAGCCTCCCCAGCAGCACCGAGCATGTGAGGGGAGCTGG
GGCCCAGGCCATGATGGGAAGCACTCTCTGCCTAAAGACTAGGGTGATGC
GCCCTCAACTGTGGGAATGAGCCCCAGCTCTGGTGTCTGCCTCGGTTTTT
CCTCCTGGACAATCAACATGAACTCCTCACCCCTCTTATCCACTTTGCAT
AAACTGAAAATAACAAACCCAGGGTCTTTCTGTCACAGGAAAGGGTTTTT
TTTTATAAGATTAAACAGAGATGATTCAACACACCCAGGATATAACACAT
GGGCCATGAGTCAAGGCCAGGCATTGCTCTGGTCAGCCTGTTGTTTGGGC
CCCCTTGGCAGGGCTCTCCCCTGAATCTTCCCCCTCTTGACTCCCCATCA
CCACAGCACGTCCAGCTTTGGGTACAAGGCCAGTAAATGGGGAAGGGGGT
CAGATGACATAAAGAGCCCTTTCCTGTCCCATTGAAATATATTTGGATAA
CAGATGGCATTTCCCCCTGTGTCTTGCCCAGGGCCCAGAGCCTCCACTTG
CTAGAGGCAGACAGAGGATGGAGAGCCCCTTCATTAGTGGGAGGACATCA
CAGGTGGGCAAGAAACCACAAGCTTGCACTGAGGCCCAGCCTTGAAATAG
CAGCACCTGCCGGCACCTGTGGTCTGGGGACAGGGTCACAGGATGGAGGG
GCCTCCTAAGCCTTTTATCTCTATGTACTAAGTACAACCCATTTTCCCAC
CTCACAGAGCCAGATCAGCCTCTGTGAGGTCCTGGTGGCAAAAGGATAAT
TGCCTGCCCGCCTGCCCGCGGTGGGGTGCTTGTGCTTGCATTCCTGGGAA
GGTTGTTGGGTTACTCTGCAATAGGTCTCTCTGACCAGCTCACCCTCCTA
CTGCAAACCTCAAACCAACTTCAAAGAAGATCCAGCACC
>Contig24
CGCGTAGTCTAAAGACTGAGTCTGAAGCTGTCCCTTCCTGCTATGGACTT
CAGATTTTAGCCCACTTGAATTGCTCCATATCCTCCAAGCCATGGCCATC
CCTTGACTCTCTGGGCTCCCAAGCACTTGCTGCCTTCATCACACAGTTTG
AGTTAAGGCAGAAAGACTGGTTTCCATGTACACTTTGTGGAAGCTTTCTC
ATTTCTTTATATAATCTCTGTCCTTTGTCTACTGCTTTAAAATCTAGAAA
TTGTTTACAAACACAAAGGTGATCCTTTAAAAGCTCAAAGCTGATTGTGT
CACCAATATATACCACTCTTAATGGCTTCCCATTAAACTTTGAGTAAAGA
CTTTATGGAGCCTACATAAGGCCATGACTACCTGGCTCTTATTTTCCTCC
TCATCCTCATCTCACCAACTCACTCTCCACTCCTATACCCCTCACTCCTT
CCCCCTCCTCTCTCTGAGCTCCAGACTCCCAATTACCTACTTCCACCCTT
TTTGACCCCCAGGGACTTATCTCAGCCTGGAATTTTCCCTCTTTGCTCTC
CACTGAACTGTCCACTCCCAGTCTAAGACATGTGCTTATGTCACACGCCC
TTACCGTGCTTATCTCAGTTTGTAATTATCTACTCATTTAGAAAAGTGTT
GATGAAGGTCTTCACTGTCAGCTTTCAGGATAGCAGGAATCATAGCTGAT
TTTACTTACTTAACGGGGTTTCATTCTTTGTAACTTTTTTTTTTTTTGAG
ATGGAGACTCACTCTTGCCCAGGCTGGAGTGCAATGGCATGATCTCGGCT
CACTGCAACCTCCACCTCCTGGGTTCAAGTGATTCTCCTGCTTCAGCCTC
CCGAGTAGCTGGGATTACAGATGCCTGTCACCACGCCCAGCTAATTTTTT
GTATTTTTTGTAAAGACGGGGTTTCATCATGTTGGCCAGGCTGGTCTCGA
TCTCCTGACCTCAGGCGATCCACCCACCTCAGCCTCCCAAAGTGCTGTGA
TTACAGGCATGAGCCACGGCACCCAGCCACTCCTTTTTTACTTATGGGTG
AGAAGCCATTAGAGATCATTTCTTCTTTTCTTTCTCTCTTCACTAAGGCA
CCAGGGTCACTAAGTAGTAGGATACTTTGAACTAGAACTCAAGAAATTGA
GTTTTAATTTTACCTCACACTCTCATATGAATTCTCCATGTGACCTCGGG
CCATACTTCCCCTGTACCCTGTTTCCTCTTTTATAAAAGTAAGAGTTTAA
ACTAGATGGTCTCCGACATGCATCCTTCTCTAACATATTCTGGAACCTTC

FIG. 3 (12 of 52)

AATAAACTAAGATAAAC AGAATAATTAAAACTTAATTTAAAAGAACA
GGAAAGGAAGCAGTTACATTAAGCAAAAGAGACATCTTCATGGTTGAAGA
AGTGTATGCCCTGGTGTCTGGATCCCATTTAGGAAACTTGGTAACCTTGC
AATCTTGGGCAGATTGCTTAATTTCTCTAGACCATGACTTCCTCTTCTGT
AAGATGTGATAAGAACATCTACCTCACAGGTTTCATGAGAGGATTAAATG
AGATAATGTATTATAATCCCTTGAACATGGTAGGCTGTTATGTTAAGTCC
TTTCCTCCTTCTCTGTAGCTATCATGGAATTTAAAAACACATTATAACTA
GAGCATGAGTTGCGACTAAAGGCTCAATTGTCTCTGCATGTGTTGGCTCA
TGCATGCTTTATTCCTCTGAAGAGCTTTTATACCAAGTGAAAGGAAATAA
TTGCATTTCCCTGAAAATTCACAGGAAAAAGTTATGTTTTTCTCTTCATT
CAAGTGATTCTGTTAGACCCAACCACATGCAACAATTTTAAAGTTGCTTC
CAAATATATTTACAAATATTTCCTGTCTTCAAGGAACAATGGCAAGACCA
TGACTCAGGTTCACATCCGGATTCCACCACTAACCATGTACCCAATTACT
TCAGTCACCTTCATTCAGGTCTTACATATCACAGAATAAAATCAGATTTC
ATCAGAGGAGGTGAAGACAGGGAGAGGAGATATTTCAATCCCTTCTCCGC
AACCCCCGTTTTTTTTTTTTTTTTAACAAGGATCCTAGAGTTACTGAATG
ATAGCACGTTTGAGGGGGAAAGACCCTAAGGATGATCTTTATAAGCCATC
ACTTGGTGTTGGTGGTGATAAAAAAACTCGAGTATCTTTATGCAGTGGAAA
GAGAAGATTGGACTCGGAATCAGAAGCTTGAGTTCAAGCACTGGTTTCAT
CAGTCTTGTGATCTTGGGTTGGTCACTTAACCTCTTCAAGGGTCCTCAGC
TGTGAAAGAAGATAGTATCAGCTAATTCTTGTATGTGCAGTGAGGAGGCA
GTGAGATAGTGCAGGTAAACTATAAAACAATTGTCACATGAAACGCATCA
CAGTGATTCTTTGGACCCACAAGCTCCAATCTTATAAAACATATCCAGTC
ACCCACCAACATAGATCATCTCACCTTGCATATCTGATTTTGTGGATCAT
GGGGAAAAACTGCTGATTCCTAGCAAAACCCATGGCATAGGATAAGTGCA
CAATAATTTTTTTTTCCTAAATGATTTAGATGACAGTGACTCATTAAGGG
TTTCCTGAGGCCTCCTCAGAGTCGAGAGGTGGGTGCCTGAAGCCACCCAA
AGTCCCTGTCACAGGATGGCTCCCAACGCACACACCACAGGCCTGCCCAG
TATGTTCCACTATCTACCCAGTAGAGCCCTGCCCAGTACGTTCCACTGTC
CCTTCCCTAGAAGAGGTGACTGTTGTTCACAGTCCCAGAAAAGCGGGCTC
CCCAAAACAATGCAAGGACCCACCTCTCTCTGAACCTCACCCACCCTAGT
TTTCCTTTAAAAATCAATTTACAAGAAGATCATGTGAAGGAAAAGGTTGG
GTGATATTCTAACCCAAGTTAGCTGTTTCTCAACCAAGTTCTCTTTGAAA
AATTCAACAACCACCTTTGGGGAATTATTTACAACAGAGGAGTGAGGATG
GGACCAGGATAGGTATTGCCTATGTTGGTGGAACCAGGGTTTTTTTCCTG
GATTACCAAAGAGATGGTATGCATTGCTCCCAGAAGCTAAATATCTTCAG
GCTTTCAATGGTGGCCTTCACCTGAAAATGTTATCCCTGTTGAAGCTTTC
AAGCCAGTATTTTCATAAGAACTATATTTTCTTTGGTGAACTGAGGCATT
ATAATGATGACTATACAGGTTCTTGAGTGACTGAAGCCATCATTAGCATT
GTCATTATTTTTGTTTAGTTGCATCTCCATAGCAGCTCACATTCACAATG
TGCTTTGCAATTGTTCCTTAGCAATAGCCCTCACAAGATTCTCAGGAGGA
GAGGGTTAATCCGGATTAACATTTCTGTGAAGCCTAGCGAGATTAATCGC

>Contig25
AAGAGTTTTAAAATTAAGTAAGGACGCCGGGAAACAAATCAATCCCAGCA
AACATTTTGTTGGGATTTATCATTCAAGCAATTTTACAGTTATCCCTGTC
AAATACATTAAGTGTTCAAAATTGGGCATAGGGGGAACAAAATAATAAAC
CCAGCCAAAACAGAATAATCCCTGTTTGTTCAATGTTGGATAAAAAAGAC
ATTACTATTGGTGTAAGGAAATTAGATACATCTTCCATTATTTAGTAAAA
TTACCATAACTTCTAACTTTGTGGCTTTAGGCAGTCTAGTCCACAGGCAG
GAAGGAGGTTTGTTTTGGCAAATGACTGTTATCATCTTCTGTTTCAAAGC
TAAACCATAAACTAAGTTCCTCCCAAAGTTAATTCAGCATATGCCCAGGA
ATGAACAAGGACAGCCTGGACGTTAGAAGCAAAATGGAGTCAGGTAGGTC
AGATCTTCTTCACTGTCTCAGTGATGGCAGTTTCATAACTTTAAATGATG
GCTATCACAGTTTTCATAAATAATCTAGATAAACAGTTAAAATAAAATAA
TTAGGTAAATGTAGTGCGATAAATATTAGTAGACAAACTCACCATAATTT
AGAATCTAAAGTTAAATTAAATAATAATATTTCATTATTTGGTATTTTCC
AAGAAAAACATATTGTAGGAAACCATTCTTTTTAAAAAAAAAAGTGTCCT
TTTAAAAAGGTGAATAATTTTTGTCTAATTCAAAGTTTATTGAAAAGTTA
TGTATAAAACAAGGTAAAGGAACAAGGAAATAAGGGAAATGTAAAGAAA

FIG. 3 (13 of 52)

ATTATAGAAATAAAGTG[illegible]TATTTTTGGTAAGAAAGCTTAAAGAGAAA[illegible]
ATTTTAGGTAAGAAAGAATCTTACCTAAAATTTTGTGCTAGAATAAAGTG
ACTGGCTAAGAAAGGGATGTTCAAAGCTATTTATGACAAACCCACAGCCA
ATATCATACTGAATGGGCAAAAGCTGGAAACATTCCCTTTGAGAACTGGC
ACAAGACAAGGATGTCCTCTCTCACCACTCCTATTCAACATAGTATCGGA
AGTTCTGGCCAGGGCAATCAAGCAAGAGAAAGAAATAAAGGGTATTCAAA
TAGGAAGAGAGGAAGTCAAATTTTCTCCGTTTGCAGATGCATGATTGCAT
ATTTAGAAAACCCCATCATTTCAGCCCCAAAACTCCTTAAGCTGATAAGC
AACTTCAGCAAAGTCTCAGGATACAAAATCAATGTGCAAAAATCACAGGC
ATTCCTATACACCAATAATAGACTAACAGAGAGCCAAATCATGAGTGAAC
TCCCATTCACAATTGCTACAAAGAGAATAAAATACCTGGGAATACAACTT
ACAATGGACATGAAAGACCTTTTCAGGGTGAACTGCAAACCACTGCTCAA
GGAAATAAGAGAGGAAACAAGCAAATGGAAAAACATTCCATGCTTATGGA
TAGGAAGAATCAATATCGTGAAAATGGCCATACTGCCCAAGTAATTTATA
GATTCAATGCTATCCCCATCAAGCTACCATTGACTTTCTTCACAGAATTA
GAAAAAACTAATAGCCAAGACAATCCTAAGCAAAAAGAACAAAGCTGGAG
GCATTGTGCTACCTGACTTCAAACTATACTACAAGGCTGCAGTAACCAAA
ACAGCATGGTACTGGTACCAAAACAGATATATAGACCAAAAGAACAGAAC
AGAGGCCTCAGATATAACACCACACATCTACAACCATCTGATCTTTGACA
AACCTAACAAAAATAAGCAATGGGGAAAATAATTCCCTATTTAATAAATG
ATGTTGGGAAAACTGGTTAGCCATATGCTGAAAACTGAAACTGGACCCCT
TCCTTACAACTTATACAAAAATCAACTCAAGATGGATTAAAGATTTAAAC
ATGGCTGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCC
GAGATGGGTGGATCATGAGGTCAGGAGATGGAGACCATCCTGACTAACAC
AGTGAAACCCTGTCTCTACTAAAAAATACAAAAAATTAGCTGGGCATGGT
GGTGGGCGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATGG
TGTGAAACCAGGAGGTGGAGCTTGCAGGGAGTGGAGATCACGCCACTGCA
CTCCAGCCTGGGCAACAGAGTAAGACTCCATCTCAAAAAAAAAAAAAAAA
AAAAAAAGAAGGATTTAAACATAAGACCTAAAACCATAAAAACCATAGAA
GAAAACCTAGGCAATACCATTCAGGACATAGGCATGAGCAAAGACTTCAT
GATTAGAACACCAAAAGCAATTGCAACAAAAGCCAATTGACAAATGGGAT
CTAATTAAACTGAAGAGCTTCTGCACAGCAAAAGAAACTATTGTCAGAGT
GAACAGGCAACCTACAGAATAGGAGAAAATTTTTTCAATCTATCCATCTG
ACAAAGGGCTAATATCCAGAATCTACAAGGAATTTAAACAAATTTGCAAG
AAAAAAAAACCCATCAA[illegible]AGTGGGCAAAAGATATGAACAGACACATCTC
AGAAGAAGACATTTATGT[illegible]GCCAACAAACATGAAAAAAAGCTCATCATCA
CTGGTCATTAGAGAAATGCAAATTGAAACCACAATGAGATACCATCTCAT
GCCAGTTAGAATGGCGATTATTAAAAAGTCAGGAAACAACAGATGCTGGA
GAGGATGTGGAGAAATAGGAATGCTTTTACACTGTTGGTGGGAGTGTCAG
TTAGTTCAACCATTGTGGAAGACAGTGTGGCAATTCCTCAAGGATCTGGA
ACCAGAAATACCATTTGACCCAGCAATCCCATTACTGGGTATATACCTAA
AGGATTAGAAATCATTCTATTGTAAAGACACATGCACATGTATGTTTATT
GCAGCACTATTCACAATAGCAAAGACTTGGAACAACCCTAATGCCCACC
AATGATAGACTGTGTAAAAAAATGTGGACGTATACCCCATGGAATACTAT
GCAGCCATAAAAAAGAATGAGTTCATTCTTTTGCACGGAACTGGATGAAG
CTGGAAGCCATCATTCTCAGCAAACTAACACAGGAACAGAAAACCAAACA
CTGCATGTTCTCACTCATAAGTGGGAGTTGAACAATGAGAACACATGGAC
ACAGGGAGGGGAATGTCACACACCAGGGCCTGTCAGGAGGTGGGGGGCAA
GGGGAGGGATAACATTAGGAAAAATACCTAATATAGATGACGGGTTAATG
GGTGCAGCAAACCACCATGGCACATGTACACCTACGTAATAAACCTCCAT
GTTCTTCACATGTATCCCAGAACGTAAAGTAAAATTTAAAAAAGAAAGAA
AGAAAGAAAAGGATGTTCACGACAAACCAGAAAGTCCAAGCATGTCATGA
ATAGTCTGTGTAAGTCACAATAAGAGGATTTATTTAAAAAAACTTTTATA
TGATAAAGTTGTCTATAATTAAAGGGAAATTATAATGGTCTTTCTAGAGA
TTGGGTTGATGTTAAAAAACTACTTATATATTAAAAAATTGGTTAGAACA
ATGAAATTTTCTTACGGGGTTGATTCACTCTTAATAAATTATAAGAGACT
TAAGAATTTTTTTTTAACCCAAAGTTCAGCTTTTATTGCATCTTGCTGTT
TTAGGTTTTCTCTCCCCTTTAAAAGGGTGGGAAATAGTAATGCCCTCCTT
CAACTCCCTTCAGCTCATATACGTTTTTTACCCTCAGATTCTGTTTGTTG
TGTCCTGATGCTAACAATGTTTTCTTAAAGGTCTAAAGGAAATGTTTTCT

FIG. 3 (14 of 52)

TCCAACATAATATTCTGTGCATTGCAGAAGGTCTTTTCTTTTGCCTTTTG
GTAACTGGCTTAACAGATTTTATGTTTTATTGAAATAATTTCTATGCCAT
TATTATTAAGTTTTGGTTTGCTTAGAAAACACTGAGATTAATACAATTTT
TTAAAAATTATGATTATTACATCCATATATCTTTATGTATGTGCTTTTAA
AGTCCTTGTGACATTGAGTTCTAGGGCTTGACTCCTGGGTCTTAAAAGGA
CAAGTCCTGCTAAATCTTAAATACTGACAGCAATTAAAGGCTCATCTTCA
GGACTGGTAGAAAATGCCAATCAAAATAAACTGCATTCTTGAAACACAGA
GCCAGAAATTAAAGCTATTCAACTCAAGGCCCAGGAACTATAGTGGAAGA
GGTGGGTGTGTGAGATTGTAAGGGCCAATTTTGAGAGATAAAATAAGTTC
AATTTCTCTATAAATTAATCATAATCATTGATGTCCAAGCCACACTGATG
CAAGATCAGCATATGGGTCCTGTGTCAGATTAACAAGGTTTTCTTGAAGC
ATTAACCTACTCCTTAATAAAGGTTATAGAGGTTATAAAAGGCTTCTGGA
AGTTATAGCTATGGTCAAGATAAAAATTTCATAGATTGTTAATACAATTT
TGGAAAACAAATTTAATTGGCTTCTTGCTGTTTTTATTAGGGCTTATTGT
TTGGAAAATTAAGTCTCGTCTCTCAAAGAATGAAGGCTTTCACCTTTTTT
TTTTTTTTTTTAATCCTTGAGTTATCACTTTGGTCAAATGAATGACTTA
TTTTACAATGACCTTTCATCAAGTGTTTTAAACCTTTCAAATTTGACAAA
CTTTCCAAAATCAAACTACAAATTATGTCTTTTTATGACCTAATGAATCC
TTTAAAATACTAGGTTCCCTAAAGTCCAAAAAAAAAAAAAAAATAACATAA
TGTGGCTTATTTGGTATAAAAATTTTACAAGAAACATTGTCAAATATAAA
ATATTGTGTGGTTTTGTTTGGGCTGTATTTGTATAAATATGTTATTGGTA
TGTGTTCCAAAATTATAGGAAACTCCTATAATTCTGATATGACTTGGTGT
ACATTATCAGTAATAATTATAATTGTTATGGTAAATTATTGTGTGCCATG
GAGGTAACAAATTTCCTCATCAAGTGTGTCTTTGACTATGGTTGCCCTAA
AACTTTTTGCCATTCACAGACAATTGTCTTGCTTTGGTCCTCTTTAGAAG
GTGGTTTTATAATCAGCTATAAAACTCTAACGGGTGCTCTTGAATGCAGG
CTTAAGATAGCTTTGGAGACTGTGACATCAGAATAGAGGAAAAACTTTCA
GTATTCATGGAGTGCTGAAATATTCATGAATATCAAGCAAAACAGGAATT
AACTTCATAGATGGAACTAAAAGAATGCTGAAGTAATCTTTTTGACTTTT
TTTCTTAGAATGTTGATCCTTCGTTTTGTTTTTCAGAGTCNAGGAAATTT
TTCTGTTGAGATATTGACAGCTTTAACAATTAAGTATACTCCAGTGAACA
CAATTTGGAGCA

>Contig26

ATCTAGTCATTCCCCAGCCTGACCAATTCAATGGCCCCCATCTTAGTTAA
AATTCCTCACCCTGACAAGGCCCCATCTACGCCTCTGACCTCATGCCCTC
CACTCTCAGTCTTGCACTCACCCTGCCACACTCAAGGGCTTCCCCAGGTT
CCTTCTTAGATTCCACCGATAGCTCAGGGACTTTGCACATGCTACGGTCT
CTGCCTGGCTCCTCCCCAGATCTTCTCATGCCTAGCTGCTTCTCATCAGC
ACCCCTCAGAGACTGTCCCTGCCCCACCTCTCCAGGTTCCATACCTGCCA
CCCTCCCCCAATCACGTAACAGTTTCTTCACAGAGCGAGTTACCATCCCA
GTATTTCCCTAACTTATTTTTTGTGACTGGTCTGTTGCCTGTCTCCACCA
CAAGAACATAAGCTGCATGTGAACAGGAGCCTTGTCTATCTTGTCACCCC
AGTGGCTGTGACATAACCTGATACACATTAGATGCTCAATGATGTTTGAT
GAATGAAGTGCTGGTAGTCCAACTGTGTTTCCTTGTCTGTGTAAGTATGT
CTGTTGTGGTTTCCTAAGAACCTACAGCTCTCCCACTGTGACTCCTGTTC
TATGGTCCTGATTTGCTGGACTAGAATCCTAACCTACATGCTTACTCTTA
GTGTCCTCCCCCAGAGGCTGAATCCCAGTCCCTAAACCTCCACCAAATGG
CTAAGACCTAGCTTCCAACCAGACAGGCCTACGCTGAGACCTCAGCACCG
CCCTTCTGCGGTCTCATCCTTAACGCATCCTTCAGGGCCCAGCTTAAATG
TCTCTTCTCCAAGGAAGGCTATCCTCTTTCTGCCCCTCAGTGCTCTCCAT
GCCTCCTCTATGCCTCCATGCCTGCTTTCAACCCTGCAGAAGTGGAGAAA
TTGCTAATCTGCTGTGTTGACACTGTGCTGGGGTGCCTTGGGCCAGGGAG
CAGGCTGGTGGTGTGCTGATAGCCCGTGGCTGTGCCCAGGTCCATGCTCA
CTTCCTGAGCCCCAGTGGAGTAGGCTCCCTTTCCCTTATTGCAGCACTCA
GAGGAAGGACGTGCTTCTTAGGACAGATCTGGCCAACCTCTCCCTCGTGA
GAGAAGGCCCAGCCATCCTCTTGCCCTCTTTCTTTCTCCTGCCCCCGAGT
AATAAAGGTGCCTGGTCAGAGCCTTCTAGAAGGAGACCCAAACATCCACC
ACACATTCCCAGTTCCAACCGTCATCCACATGGCTGGCTGTGCAGGTAAA
CGCAGAGTCTGTTTCACACACCCAACCATCTAGTATTGGATGGGAGGACA
GTAGCGTGACACTCTTCTCCAGCCTTGAGCCCTACTGTGGGCCCCACCCA

FIG. 3 (15 of 52)

ACCCAGATACCAGAGGAGCCCTGTACTGGGATGCTATTGGATGCTTGTCC
AGTCATGTACAAAGTTAGCCCTTTGTTATATAGAGTTAGCTACGTACATC
TTCCTCTGTAGGGAACCCAAGAGGGGAGAAGAGATATGTAGTAGGATTTA
ACCTGCAAATCCTCTGCTGAGCACCCTGCACTACATACAGTGGGTAGCAT
GTGGTAGGTGCTCAATAACTATTGACCGATAGATTGAATACAGGTAGGAT
GGTGACACAATCTAAGATCCCAGGGGTGGGGAGACCACACGCTTGGTTAG
GGAGACCCAAAGTGGACCGTGTGGCCAGAAGAGTCCCGCACTGCACTCTA
GTGACAGTGCAGAAAGTCACTGTGGGAAATCTAGAAGTTTCTACAGGTTG
CTATTTCATCATAGCACTGTGCAGGCCAACCCTTCCTGCTCCACTGGCTG
TTGGGAAAAGCTTTCTCTTTTCTTCCTAGCCAGGGAGCTCTCAAAGTGTT
CCACTCTCTCACCTCCACCCAGGCGTCCAGGTGTGGAGGACACTTGCCGG
CTGCTTGTCTGCTGACTCATCCCTTGGTTTCACTTGGAAAACCTACCACC
AGCTGGCCTCTTTCCAAGCATCAGCCTCCTCATTTTCTTAATCCCTTAGG
TGTGATCTCACCTCCACACAGTAGATTGCCTCAAGGCCCAATTCCAATAT
GAATAAAAATGATTATTTTGTCATCTTCCAATCTTCCTTTTAAAATATTA
TTTTATAATTCCCTTTAGGAGGATCACCTAAGTGAAGACTATTTTTACCT
AAGAAATGTTAAAATGTAAAGACATGGTTGTAATCTGGGGATTCCTGTTA
AAATGGCTAGCAGACAGAAGTCAGACGACAGGCTAGAAATGTGTGAAGAG
TGGTTGCCTTTGAAAGGCGGAGTTGGTAATGATTTTCTTCCATTTTTCCA
TGCTTTCCAATTCTCTACAAAGGCCTTAATATTACTTCGATAACCAGGAC
CTCTGATAACCTGCCCCCACCGAGTAAAGACTTAGCTGGGAAAGTCAGCT
TCATGTGAGGTAAAAGGAACCAGGTAATACACAATTCCCACTGCCAACTG
TCGGGTGTGCAGGCCTGAGCTTCCTGCATGTGGGAGGAAAGAGAAAGAAG
AGAGAAACTCCAAGATCCAAGAGATCCAGCAAGAAGGCTGGAGTCTGAGG
ACGCAGAAAGCTGAATGGCACAGTTACCACTATTGTGCTGAGGTTCTGTG
GCCTCTGGGTCTCTTGACAACTGGGCAAAGACCCACAGAAAACTATCTCT
AGACCCTACCTGTGGGAGGGGAAAGTGCTTAAGATCATTTACAGGACAGC
CACCTGGACCTCAAATGGCTTACAGTTCCTTCATCCAGAGGGTCTTCATT
TAGTACATACCAGGTGCTAAGCTGGGTGCTGGAGACATGACGGGGAACCC
ATTTACCATGGCTTTGTTACTGTGACATTCACATCTAGGGAAAGCCAGCA
AAGGGGAGGGATCGAGGAGAGCTTGTTAGGCAGAGAAAATACCCAAGGGC
AAGGGAGAAGCCAGCCTGTTCTGAGCACACACAGTGGTTCCATCTAACTG
GGCCTCAGTGCCAGGTTGGACTGGAGATGGGGCTGAGGAGCTGTCACAGA
GCATTCTGGACACAGATGTCACATAGTCCCTTGAGGTTAGGGTCCTTAGG
CATGGCAGCATTGCTTTGAGTTTTTCCTTTTGTAATGTTGCCATTCATGA
CAATGTGGAAGATGGGTCCTTGCAGAGAAGGGCAGGGCTGTGAGACCAGT
TAGGAGACTAAGATGTGAGCCAAGGAAAATGAGGAACACCTGAACACTGG
GGCAGGTGCAGGGCCCAGAGAGAAGCAGATGGCTTCCTGAGGTTTTAAGT
AGGTAGAATCAAGGCAGCTGGTACAGATCTTTTATTACATATAAACTGGA
ATAAGCCATCTGTTCCAAGACAAAAGAGTAGGCGGAAAACAATACAAGAC
AGAAATGGAATTAGAACAAACCTGGGAGGAATGTGGAATTAGAGTAGAGA
GTCCAACACTGGCTGCAATCATAAAAATGTAAAACAAACAAAAATTTGCT
AGGTGTGCTTACTTAGAAATAATTAGCTGTCATATTAAGTTCACTTGTGT
TATGGCTTAAATGTGTCCCCCAAAATGTGATGTGTTGGAAACTTGATCCC
CAATGCAACAGAGTTGAGAGATGGGACCTTTAAAAGGTGATTAGGTCATA
AGGGTTCTGCCCTCATAAATGAATTAATACTGTTATCATGAGAGTAGATT
CCTGATAAAAGGATGATCTCTGCCTCCTCCCCACAGCCCTCTTGTGCATG
CTTTCCTGCCTTTCCACCTTCTGCTATGGGATGACACAGCAAGAAGGCCC
TCACCAAATGCAGCTCCTTGATCTTGGACTTTCCAGCCTCCAAAACTGTA
AGCCAAACAAATTTCTGTTTATTATAAATTACCCAGTCTCAGGTATTCTG
TTCTAGAAACACAAAATGGACTAAGATCATTAAATTATCATTTTTTATCA
GACTGTTGA

>Contig27

AAAATATAACAGAGAGTAAGAGGAAAATTACCTTCTTTCTTTTTCCTTTC
CCTGCCTGACCTTATTCACCTCCCATCCCAGAGCATCCATTTATTCCATT
GATCTTTACTGACATCTATTATCTGACCTACACAATACTAGACATTAGGA
CAATGTGGCCTGCCTCCAAGAAACTCAAATAAGCCAACTGAGATCAGAGA
GGATTAATCACCTGCCAATGGGCACAAAGCAACAAGCTGGGAGCCAAGTC
CCAAAATGGGGCCTGCTGCTTCCAGTTCCCCTCTCTCTGCATTGATGTCA
GCATTATCCTTCGTCCCAGTCCTGTCTCCACTACCACTTTCCCCCTCAAA

FIG. 3 (16 of 52)

CACACACACACACAACAGCCTTAGATGTTTTCTCCACTGATAAGTAGGTG
ACTCAATTTGTAAGTATATAATCCAAGACCTTCTATTCCCAAGTAGAATT
TATGTGCCTGCCTGTGCTTTTCTACCTGGATCAAGTGATGTCTACAGAGT
AGGGCAGTAGCTTCATTCATGAACTCATTCAACAAGCATTATTCACTGAG
AGCCTTGTATTTTTCAGGCATAGTGCCAACAGCAGTGTGGACAGTGGTGC
ATCAAAGCCTCTAGTCTCATAGAACTTAGTCTTCTGGAGGATATGGAAAA
CAGACAACCCAAACAACCAACAAAAGAGCAAGATGCTGCAAAAAAAAAAA
AAATGAATAGGGTGCTAAGATAGAGAAAAGTGGGAGAGTGCTATTTAGAC
AAAGTGGTAAAAACAAAGCCCCTTGTGAGATGAGAGCTGCCGACAGGAGG
GGGCGGGTCATGGTTGTGGGTTTTTGGGTAGGACATTCAGAGGAGGGGGC
GGGTCGTGGTTGTGGGTTTTTGGGTAGGACATTCAGAGGAGGGGGCGGGT
CGTGGTTGTGGGTTTTTGGGTAGGACATTCAGAGGAGGGGGCGGGTCGTG
GTTGTGGGTTTTTGGGACATTCAAAAGAGTCTGAATGCACCCAGGCCTAC
AACTTCAAGATGGTAAAGGACAGCTCCAAGGATCAGAAGAAGCATGCTTG
GAACTGGGGCATTTTGAGAAGGAGGAAAAATATGCAGAGACTAGTGCTTG
CAGAGCTTGCATGTGGATTTCATTTGAGGTACAATGAAAACCCATTAATG
GGTTTCACACAGTGCAATGGCCTGACCTCACTTATATTTCCTAAAATAGA
AAACAGATCAGAAGGAAGGCAATAGAGAAGCAGAAAGTCCAATGAGGAGG
TTTCACAGCAGTCATGGGGGTGGGGTAAGGAAAAGAAGTGGAAAGAAACA
GACAGAATTGGGTTATATTTTGGAGATAGAACCAACAGAAGGAAGAGGAG
AAACAACATTTACTGAGAAGGGAAAAAGTAGGAGAGGAATAGGTTTGGGA
AATAAATCCTGCTGACATTGGAAACCCCAAGGAAGCCTCAAAAGTATATT
TACTTGCTTTAGATTTAAAAGAATAGGAAAGAAGCATCTCAACTTGGAAT
TTGAAATCTATTTTTCCATAAAAGTATTGTTAAATTCTACTCATACTCAC
AAGAAAAGTACATTCTAAAGAGTATATTGAAAGAGTTTACTGATATACTT
AGGAATTTTGTGTGTATGTGTGTGTGTGTATGCGTGTGTGTGTGTTTAAC
CTTCAATTGTTGACTTAAATACTGAGATAAATGTCATCTAAATGCTAAAT
TGATTTCCCAAAGGTATGATTTGTTCACTTGGAGATCAAAATGTTTAGGG
GGCTTAGAATCACTGTAGTGCTCAGATTTGATGCAAAATGTCTTAGGCCT
ATGTTGAAGGCAGGACAGAAACAATGTTTCCCTCCTACCTGCCTGGATAC
AGTAAGATACTAGTGTCACTGACAATCTTCATAACTAATTTAGATCTCTC
TCCAATCAACTAAGGAAATCAACTCTTATTAATAGACTGGGCCACACATC
TACTAGGCATGTAATAAATGCTTGCTGAATGAACAAATGAATGAAGAGCC
TATAGCATCATGTTACAGCCATAGTCCTAAAGTGCTGTTTCTCATGAAGG
CCAAATGCTAAGGGATTGAGCTTCAGTCCTTTTTCTAACATCTTGTTCTC
TAACAGAATTCTCTTCTTTTCTTCATAGGAGATGCCTGAGATACCCAAAA
CCATCACAGGTAGTGAGACCAACCTCCTCTTCTTCTGGGAAACTCACGGC
ACTAAGAACTATTTCACATCAGTTGCCCATCCAAACTTGTTTATTGCCAC
AAAGCAAGACTACTGGGTGTGCTTGGCAGGGGGGCCACCCTCTATCACTG
ACTTTCAGATACTGGAAAACCAGGCGTAGGTCTGGAGTCTCACTTGTCTC
ACTTGTGCAGTGTTGACAGTTCATATGTACCATGTACATGAAGAAGCTAA
ATCCTTTACTGTTAGTCATTTGCTGAGCATGTANTGAGCCTTGTAATTCT
AAATGAATGTTTACACTCTTTGTAAGAGTGGAACCAACACTAACATATAA
TGTTGTTATTTAAAGAACACCCTATATTTTGCATAGTACCAATCATTTTA
ATTATTATTCTTCATAACAATTTTAGGAGGACCAGAGCTACTGACTATGG
CTACCAAAAAGACTCTACCCATATTACAGATGGGCAAATTAAGGCATAAG
AAAACTAAGAAATATGCACAATAGCAGTTGAAACAAGAAGCCACAGACCT
AGGATTTCATGATTTCATTTCAACTGTTTGCCTTCTACTTTTAAGTTGCT
GATGAACTCTTAATCAAATAGCATAAGTTTCTGGGACCTCAGTTTTATCA
TTTTCAAAATGGAGGGAATAATACCTAAGCCTTCCTGCCGCAACAGTTTT
TTATGCTAATCAGGGAGGTCATTTTGGTAAAATACTTCTTGAAGCCGAGC
CTCAAGATGAAGGCAAAGCACGAAATGTTATTTTTTAATTATTATTTATA
TATGTATTTATAAATATATTTAAGATAATTATAATATACTATATTTATGG
GAACCCCTTCATCCTCTGAGTGTGACCAGGCATCCTCCACAATAGCAGAC
AGTGTTTTCTGGGATAAGTAAGTTTGATTTCATTAATACAGGGCATTTTG
GTCCAAGTTGTGCTTATCCCATAGCCAGGAAACTCTGCATTCTAGTACTT
GGGAGACCTGTAATCATATAATAAATGTACATTAATTACCTTGAGCCAGT
AATTGGTCCGATCTTTGACTCTTTTGCCATTAAACTTACCTGGGCATTCT
TGTTTCATTCAATTCCACCTGCAATCAAGTCCTACAAGCTAAAATTAGAT
GAACTCAACTTTGACAACCATGAGACCACTGTTATCAAAACTTTCTTTTC

FIG. 3 (17 of 52)

TGGAATGTAATCAATG1.TCTTCTAGGTTCTAAAAATTGTGATCAGACCA
TAATGTTACATTATTATCAACAATAGTGATTGATAGAGTGTTATCAGTCA
TAACTAAATAAAGCTTGCAACAAAATTCTCTGACACATAGTTATTCATTG
CCTTAATCATTATTTTACTGCATGGTAATTAGGGACAAATGGTAAATGTT
TACATAAATAATTGTATTTAGTGTTACTTTATAAAATCAAACCAAGATTT
TATATTTTTTCTCCTCTTTGTTAGCTGCCAGTATGCATAAATGGCATTA
AGAATGATAATATTTCCGGGTTCACTTAAAGCTCACATTACACATACACA
AAACATGTGTTCCCATCTTTATACAAACTCACACATACAGAGCTACATTA
AAAACAACTAATAGGCCAGGCACGGTGGCTCAGACCTGTAATCCCAGCAC
TTTGGGAGGCCAAGGTGGGAAGATCACTTGAGGTCAGGAGTTCAAGACCA
GCCTAGGCAACATAGTGAGATCTCATCTCTACAAAAAAAAAATGAAAAAT
TAAAAAATGAGCTGGACATGGTAGTACACACCTGTAGTCCCAGCTACTCG
GGAGGCTTGAGGTGGGAGGATCACTTGAGCCTGGGAGATGGAGGCTGCAG
TGAGCCATAATCACACCATTGCACCCCAACCTGGGCAACAGAGTGAGACC
CAGTCTCAAAAGATAAATTTTTAAAAATGTTAAAAAATATATAAAAGAGA
ATTTTAAAAGAACAACTAATAGATCAAAGCATGGATGCAAGATATATTTA
GTTGGAAAATCAAGGTTAAAATCAAGGGATCTTGGAATTAGGTGTGGTAG
ATTTGGGTAAGGAGTAGTCTAAGATGACCCTGTTTCTTGGTACTGGAGAC
TGGATGAGTGGCAGCGTCTTAACCATATTTTTGGTAGAAATATGGAGGTC
TTCTCCATTCCAGGATGAATGATGAGTAAAATTTTAGGCATGTAATTTGA
GCTACTAGAAGGACACTCAATTGCAGATGTACAATGGGGAGATGATAACC
TATCTGGAACTCAGAAAAATAACTGTATATAGATATGAAAGACATCAGTA
GGTATGTAGTAGATAAAATCCTAAAAGTGATGTCAAAGGGAGAAGAGAAG
TATATGGTGAACACTGTTGTTTGTCCATGCAATTGCCATCTCTTCTTCTT
CCTTACTGACAGAACCCTGATTTCACTGAGAAGTCAACATGCCCTTCCCC
AATTGATGAATCCAATTGGTTGAAGATTATGTTCATTCTATTCTTACATG
ACTAAGTCACGTTGACTTAATCCTATCAAATGAGATGTCGATCTGGAAAC
AACTTCTGGAAAAGATTTTCTACCTTGATAAAATAAAGAGCCATATAGAT
GGTCCTTTATCTTCCTTCTTCCTTGAATGAGATATGTTCTATGAGGAAGT
GAAGCTTAGAACTGTGGTCAGCAACTTGCAACGACTGGGAAGTCAGAGCC
ACACAATGAAGAATGCAGAGTGGAAGGAGAAAAAGAGCCAGCATCTCTGA
CAACATTGTTACACCGAGAACCTACCTCCAGATTTTAAGAAAACAAGAAA
TGCTACTGTTATTAAGCCATTTCACTGGGTTTGCTATGACTTGCAGTCAA
ATCTAGCTTAACTGATACAGAGCACCACAGAGAACTGGTCTCTCATTTGT
CTCATCCTGTTCTTTCTAGCAGCCACGACTTTCCTAGGGTTTCCTTAGCC
CAAGTCTGGCTAGAGCAAGACTAAGTAAGACTTGATTCCTTAATGTCCTT
TTGTTTTAAGAAATATTAAAGAATTATTTTTATATTAATATATTTTAAGA
AATAAGGAAATACAAAACACTGAGCAAGCAACACAAATTCAAGAAATCTT
AAAAAGTATAATAGCTGCTCAGTCTCTGATTAACAGTGAAATATGGAATC
ATTGTAGAAATGGCCTTGGAGCGTTATTCTCCCAGGCCAGCTATCCTTAT
GGTCTGCCCCACCTCCCTCATTGCCTAAACAGTAAGAGAGTCCCATGGTG
AGACTCAACAGTCTTAGCACAGAACTTGTTACAGTCTATTTCTTTTCTTA
CAGTCCTATATATCAATTCCAAATCAATGAGAGTAAAGCCCAATCCCTGC
CTTTAAACCCAAAGGACAGAAGCCCAAAGCCCAAAGATATTCCCTAACCT
TCTCCCCCT

>Contig28

CCTGTCGCTCCCTATGTTTAAAGCTGGGGATCTCTTTTTCCTGTGTCTAA
TTATTTTCCTCATTGGCTTGAAAAATCTGATAAAACATTTTAGGACTGTG
TATAAAATAGAATTAGCCAAGTGCAATGTCTTTATTCAGAAGAAATTTCA
TGGACGTTGTGCCTACTCTCTTGGCTTCCTGGCTTCATGGCTTTCCAGAT
CCCACAGTAAGCTCTGGATAGTAGAAGTTATAGTAAGACTGACTTCTAAA
TAAATGAAGTGACTTTAACCTTACTGATATGGCTTAAAGAAAAGGAGTGG
CCTTTAAGATCCATGAACTTCTCAAACAAAAGTGATAACGTTATCTCCAT
GCATATATAATACTAAATATAATGCAACTGAGAGAAGTAGGCTGTGGTAA
GAAAGGAGACCCAAGTGCCATCTGAAGGCAGCACTTACCACTCTGCTTCA
TCCCACCGAGGAAACAAAGCATGAGTATTGCCAGATTTTCTTCTGTTTCA
AGAAAAGCCAGAAATCCAGGTTTTTGCGTGAAATGTCCTGATTTTAATGT
TGGGAACTAATTTATATTTTGAAATAACATTGTGTGGGACAAGTGAACTT
GTATGTGGAACTGCTTTCTCCCAGTGGCGACCAGTTTGGACCGTTGATAC
TCAGCAAGTTCAGCCAAGTGCGCCTTGTCATTGTCAGTCATCAAGGTGAT

GTGTGATTGGTCAAACAATTAGTTTTGCTCAGCATCTCGTGTGTTTTCAA
AGGACCTGAGGGTTCATTTGCCCATGCAGATCTTGTAGTCCTGTTTATTC
TATTAATTTATCTTGCAAATCTATAATGTTTTATTTTAAGCAGCGAGAGC
CGTGGCAGCCTTTGGTCTGGACCCTTTCTAATGATCATTTAGTATCAGGC
TATGTGGGAGTTGATTGTTTTGCATTGCCTGAAAGCCAACAGTATCACTC
CTCCTCTAGGTGTGGCAGAGATGTGAGAGAGGGAGACTGACAGTCTGTGG
GTGTGTATGCAGTGTTGGGGGAAGCGAGGCACAGGGGACAATACTGTGGT
GTATAAAACTAGTCTAAGGTAGCATCAGGAAGTTCATGAAGCCAAAATGA
TTTTCATAACAGCACAAGACATTATTTGTTTTTGCCTCCCCTCTCATTTTT
TTTTTTTTTTGAGACAGAGTCTTGCTCTGTCATCCATGCTCGTGTGCAGT
GGTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGGGTTCAAGCAATTC
TCATGCCTCAGCCTCCTGAGTAGCTGATTACAGGTCTGCACCACCCCGCC
GGCTAGTTTTTGTATTTTTAGTAGAGATGGGGTTTTGTAATGTTGGCCAG
GCTGCCCTGTCATTTTTTTTTTACTAGTGTCCAGTGGAGTTTTTTAGGGG
CTACATAACATGATACTGTCATTAATCTAATGGCTAATGAAAGGGATATG
TATATGTTTTTGTGTTTAAAACAAACTTCTTTGGGGTCCTCAATAATTTT
TAAGAGTATAAAGGGGTCCTGAGATCAAAGAGTTTGAGTTCTGCTGGACT
GGGACAGTGGTTGTCAACCCAGATTGTACATTAGGGTCATCTGGGAAGCT
TTAAAATAGTACTGATGCCCAACCTTACCGCAAACCAATTAAGCCAGAAT
CTCTGTGGATGAGAAGTCTTCATTGTCATCATCACCATGACCATCATCAT
TGTCACCGTCACTACACCATTATCATCATCATCATATCATCTTCATTATC
ATTGTTAGTATCTCCATCACCATCATCAGCATCACCATTATTATCATCAT
CATCATCCCCACCATCATCCTCATCGGAACTTCACCTGCATGGAGGACAA
TCCACTATGCATTAGGTGCTATGCTATTTGCTATACTCCTTATTCTCACA
ACTGCCCAGAGAGGCTGATATTATCTCACTTTATAACAGGAGGAATCTGG
ATCGGAAAAGTTAAGGTAAGCTAATTCACAGAGCGAGAAGAGATAGAGCC
AGGATTCGAAACCAGTTCTCTGCTACATCAATGTTCCCAGTCCTTGCACT
ATTGAGAACCTCTTTAGTTATGCTTTCACCCCTCCAACACCACAGTAAAT
TTTTTCTTTTTTTAAAAAAATTATACTTTAAGTTATAGGGTATATGTGCA
TAATGTGCAGGTTTGTTACATATGTATACATGTGCCATGTTGGTGTGCTG
CACTCATTAACTCGTCATTTACATTAGGTATATCTTCTAATGCTATCCCT
CCCCGCTCTCCCCACCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACC
CTGTGTCCAAGTGTTCTCATTGTTCAGTTCCCACCTATGAGTGAGAACAT
GTGGTGTTTGGTTTTCTGTCCTTGTGATAGTTTGCTCAGAATGATGGTTT
CCAGCTTCATCCACGTCCCTACAAAGGATATGAACTCATCCTTTTTTATG
GCTGCATAGTATTCCATGGTGTATGTGTGCCACATTTTCTTAATCCAGTC
TATCATTGCTGGACATTTGGGTTGGTTCCAAGTCTTTGCTATTGTGAATA
GTGCCACAGTGAACATTCATGTGCATGTGTCTTTATAGCAGCATGATTTA
TAATCCTTTGGGTATATACCCAGTAATGGGATGGCTGGGTCAAATGGTAT
TTCTAGTTCTAGATCCTTGAGGAATTGCCACACTGTCTACCACAATGGTT
GAATTAGTTTATAGCCCCACCAACAGTGTAAAAGCATTCCTATTTCTCCA
CATCCTCTCCAGCACCTGTTGTTTCGTGACTTTTTAGTGATTGCCATTCT
AACTGGCACCACAGTAAATTTTTATAGATTTTATAAGCAAATTGTATTTA
CTGTGCAAGAATTGGTTTATTTTTTAAACCATGTGTTGCAAACATACAAT
GGTTAATTGTGATATTTGCTCAGTACAAGATCATCAGATCACTACACAGA
CTTGAGGTAATTCCACCTAAAAGCAAAGAGAACTGACCCCACATTAACTG
AGAAGTCTTTACTTATTTATTCCCTATAAACGAGCCAATATGAAGAGAAG
GCCTTAATGTGGTTAACTATGTAATTTTTTTCTGACTTTTTGAAATACTG
AGAAGAGCTCATGACTCTCCCATCTCCTAATTCTACCTTGGTGGATTTTA
GACTGACCACAACTCATGGGTAAATGAGGGAAGACGAATAAGAAACCTTG
CTTTTTTTCCTCCTTGTTTTTGGCTGGCTGCAGTGGCTCACACCTGTAA
TCTCATCACTTTGGGAGGCCAAGGTGGGAAGATCACTTGAGCTCAGGATT
TCAAAACTGGCCTGGGCAACATAGTGAGACCCCATCTCTAAAAAAAAAAA
AAAAAAAAAAAAAGGCGACAGGCGGTGCGTGCCTGTAATCCTACCTACTC
AAGAAGCCGAGGTGGAAAGATCACTTGAGCATGGAGGTCAAAGCTGCAG
TGAACCTTGATTGCACCACTTCATTCCAGCCTGGGTGACAAAGCAGGACG
CTGCCTCAAGAAAACAAAAACAAAACCTTAATTTTTTGGCTATTCTTTTC
TGGTAAGAATGGTATAGAGATGGGGATGAGGATGGCTATTGTATGAGAGA
GCAAACAGGGTCCAAGCAGTGCTCTGGGCTGTCTAAGGACCAGTAGTCAG
CTTAACTTCTCAAATTTCCAGGGAAGGAGTTCGGAGTGGTAGAATATCCT

FIG. 3 (19 of 52)

GGGTATGCCCAAAGCATCACCTTGCAAATAGCCTGTCATGAATAATTTGT
TTCATTTGTTATGACTGGAAACTGGCTTTGTGTATGCCAGAGAATGGGG
CAGGAAAGAGAGATTGGTGTCTTGAGCTCTCTGTGCCTCTGGGGCAGTGA
TGCTTTTCCTCTCATGTGGAAGGAGAGCATGACTGAAAAGGTGCACAAAT
AAGGTGTCTGTGAGAGAAATTAACCTTCCAGATACAGAGACACAACCTTC
CCCAAGAGGTCCTCATTGCTCTGCCTTTTTTCCTTTTTTTTGCTTGTTCT
ACCATTAATAACAGAAACTGATTATGACCTCAAAAGAGAGGAGAAAGCGA
CTCTCCCCACCCTAGAGCTAGTTAACCACCATATCTTCCTAGATATCCTT
GAGAGCAATGTAACCC
>Contig29
GTGAACTCGTTTTACCTGTGTAGCAGACCAAGCCGCAGACAAAATCCNTC
AGACACCAAATTAAAGAAGGAAGGGCTTTATTGGGCCTGGAGCTGCGGCA
AGACTCACGTCTCCAACAACCGAGCTCCCCGAGTGTGCAATTCCTGTCCC
TTTTAAGGGCTCACAACTCTAAGGCGGTCCACATGAGAGAGTCGTGATAG
ATTGAGCAAGCAGGGGGTATGTGACTGGGGGCTGCATGCACCTGTAGTTA
GAATGGAACAGAACATGACAGGGATCTTCACAGTGCTTTTCTTATGCAAA
TAACCGATTAGATCAGGGGTCGATCTTTACCAGGCCCAGGGTGTGTCACC
GGGCTGTCTGCTTGTGGATTTCATTTCTGCCTTTTAGTTATTACTTCTTT
CTTTGGAGGCAGAAATTGGGCATAAGACAATATGAGGGGTGGTCTCCTCT
CTTACCTGCGGGGAGTGAGCTCAAACTCCTTAAAGGAGTTACCTGCCTTC
CATCATCAGGGAAGCAGGAAATCTTGCCTTCCTTGTTGGAAGCAAGTAAA
ACTCAAAACAAACAAAGAAAAAAACAGGGAGTTGTACAGCAAAATAAACT
TTTGATTTTGACCAAATTTTGGGAGATCAGGAATTCTCTGAAGGAGATGC
TTTCAGACCTCAGCAAATTGTCCTGTTGGTTTGAGCCATAAAGTTAGCTC
ATGCTGGTACCAAACACCAGTAGGAGATTTGTCAAAGGTAAGAGGCATCT
CCACTCAGAATCCCTTCGTGGTTACCAACATGTGAACCTTGGAAATCTGA
GACAGGTCTCAGTTAATTTAGAAAGTTTATTTTGCCACGGTTGAGGACAC
CCACCCATGACAGAGCATCAGGAGGTCCTGACCACATGTGCTCAGGGTGG
TCTGAGCACAGCTTGGTTTTACACATTTTAGGGAGACATGAGACATCAGT
GAATATATGTAAGATGTACACTGGTTCCCTCCCAGAAAGGCAGAACAACTT
GAAGCAGGGAGGGAGCTTCCAGGTCACAGGTAGGTGAGAGACAAACAATT
GCATTCTTCTGAGTGTCTGATTAGCCTTTCCAAAGGAGGCAATCAGATAT
GCATTTATCACAGTGAGCAGAGGGGTGACTTTGAATAGAATGGGAGGCAG
GTTTGCCCTAAGCAGTTCCCAGCTTGACTTTTCCCTTTAGCTTAGTGATT
TGGAGGCCCCAAGATTTATTTTCCTTCTACATCACTGTGGGCAGCTGACT
AGGAAAGCTTTGTAGGACTGGTGGGCAGTGTGAGAGCCCAGTGGGGGGTG
GTGGTCCTGTGCCAATGGTAGCAACCACCTGTGAGGCTGAGTAAACTCAT
TTCCCAACCTCCTCTAGCAGCCCCAGTGGAGATACAGAGGAAGCAGACTA
GCGATACAACCCAGCCTGAAGTTTTGTCTGGTGAGTGTAATGGAATAAAA
ATGGGAAGGGTGCTGAAGAGACCAGCAAGAAAATGGTTGAAGAGATGGGG
CACAGAAATTAAGCTGGATCAAAAAGGACGGAAAAGCAGAAAGGGCCGAT
AGAGAGAGGGGATATCTATGGGTTCGCGATTCTGAAAAGGACAAATCACT
GGTGCTTTGAGAAGAGAGAGGGTGAGAAAGCAGGAAGGCTGGAGGCTGTC
ATCCAAGAGGCGGACATCTGTGAACATGATTCCAAGAGTCACCAGACCAT
GGGGGTGGCCAAAGGGAGTGCCTCTTCTCACCTCCTACTCTTAATTCCTT
GTACTCAAGATAATAAGTTCCCAGAAGAGAAGTACCCATATTTAATTCAT
CTGTGTCTTCCTAGCAGTACTAAAAATATTATATGAAAGGTATCAAACCT
TTGAGAATGTGTGCTGCTAAATTGTTAAGGATGCTGGAAAACTCAAGACG
TCCCTGATCCTGAGCCTGAGTATGAGCCTGTGGTGAGCCCAATGCAGGTC
TCCATTCAGACAAAGGCCTCAGGGAACGGATGAGACCTAGGGACAGAGAT
GCATGCTGGAGCAGCATTCCCCATCCCTACTGCAGCTCAGGCCAGCTGAC
TGCTTTATGAGTAAACGTTACCAGGGAACACTTTGCAGTCTTAACACACA
TGCCCACCTGTGACCACTGATCCCTGTTGGGTGACCACTGACATCAGAGA
TTCGATGGCAGCAATGAAGACAAGGCTATCCTCATTAGGAAGGAAAGGAA
GGAGGAGGGAGGAGGGCAAACGAATCTTTCCTGCTTGTCAACCACGTCCA
TCTCTGTTAGGTGATTTCCCATGTGTGACTTTGTTTATCTTTATAATAAC
TCTGAGAGGTAGGTCTTGATGTCCACATTTTGAACATGAGGACATCCAGC
CAGGAAGTTGAGTTCTGGGGACATAGCTGAGAGGGCAAAGCTACATATAA
ACCCCTCTTTGTTTTTTCTGGCTTATCCACTGAGTGCCCCCTGCAATCCA
CCAGCCCATTTGTGAAGTGCATACTATAGGTAAGTTGGCACAGGAGGAGT

FIG. 3 (20 of 52)

GGATGTGGGCGATTTTG.CACAGCTCTCCAGGAACTTACACACTGGTGAG
GAGGGCCAGGTATGTTCCTGACCAGTCACAATCAAAGCAACCTCCTACTA
ATCAGGGAGGCTTGGTACCTGGGGAATGCTATGTTGAAAGGTTCTTTTCT
GGGTTTTAAAATGATGGGTCTATTTCCTTATTCTTAAGATTGCTTTTTTT
CTGGCTAGAACTTAAAAGAAATTTTCAGTAAAATTTCCCTTCCCTGGCAC
AAAGTGAGCTTGAAATGAATTCCCAGGTGGCCTTGATACTTTAAAATATT
GCCTCCTATAAAATCAACCTTTAGAAGAAGGAAGTCAAAGAACATGCTAG
ATTTCACAAAGGTTAATTCCTTGAAATCCAGTTATCTACAGGACAATGTT
GTCAAAGAAAAAATTATTTGGCCAGGCACGGCGGCTCATGCCTATAATCC
CAGCACTTTGGGAGGCTGAGGCAGGTGATCACCTGAGGTCAGGAGTTCGA
GACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA
AAAATTAGCCAGGTGTGGTGGTGGGCACCTGTAATCCCAGCTACACGGGA
GGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGAGGAAGTTGCAGTGAG
CCAAGTTCAAGCCACTGCACCCCAGCCTGGGCAACAGAGCAAGACTTTGT
CTCCAAAAAAAAAAAAAATTCAATGATATTTTTAAATTCATGGTAAGGAA
GATTTCATTCAGAACCAGCACAGAAGATATAGGAAACACTGCAATGGGAC
TTTGCGGTGGGGGAGAGAGATTGAACACAACTACATATACAGCACGGGCA
AGGACATATTCATAGCCAGGAAGCAGAGCAAAGATCAGTGGATGCGAAAT
TACTAAGAGGAAACATGAAAAATAAGGGAGCTTCTGCCTAAACCCACCTA
ACCGGATCCTTGCTGAAGACAGGACAGGGTGATTGGACACCACTTTGGGG
ATGGTGGAGGATGGGGAATCCAGTGAGATTTCAAGGGTGATGCGATATTG
AACATACAAAGTTCTTGCTAAAAAAGGATTTTACAAGAAAGTGTACAAAT
GTGCCTGGGACAAGGTGCAGGAGCCCGACGGAGATGTGGTCCAGCAGAGA
ATATGTGCCGAGATGATAGGTGAGTTCTCTGACGAAGGATATATGCTGAT
CCAGCCAGGGTGAAATGCTCAGAGAAAGCACGGAGGGGCTATGTCCGTTG
CCCCAGTCTCCACGCGGTCAAATCTGATCCCGTTGTGAGTGTGGCCGTTT
GTAGAAAGCAATCAGGGGGGGTCCCTCCCC
>Contig30
AATATATATTTTTTATANNATNTGAGACAGGTTCTCACTAGGTTGCCCAG
GCTGGTCTTGAATTCCTGCCTTCAAGTGACTCTCCCACCTTAGCCTACTG
CATAGCTGGGATTACAGGCACAAACCACTGCATGCAGCTAACTTTGCTTC
TCATTCCAGCACTTTTTATTCCACTGATTATATGTATATGTATATCTGCA
TCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATATATATATATATAT
ATGGAAATATCTCTCTCTCTCTCTATATATATATATGGAAATATATATCT
CAGTCTCTCCTATCCTCCTTTAATCAGTTTTGCTATCCTGTCAATTCCCC
CAACGAGTGTGATGTTGTGAAATATATATTTGTTCTTCATCTCCTGTTTC
CTGACATACAGCTTTTAAAAACCCTTGGAATCTCTGGAATAATAAGAGTG
TCTTTTGCATGCTAATAGATGACTGCTGGCTGGCAGCCCCAATGCAGTAG
CTTCATGATGGGGTTTGTCACAGGAAAGACCAAGGCAGGATTGGAGACTT
GAGACTGTTAGCCCCACTCCCCAACCACTGGAGGGAGTGGAGGGGCTGAA
GGTTGTGTCAGTCACCAATGGCCAATGGTTCGGTCAATCATGTGTATGTA
ATAAAGCCACTCTTAAAAACCCAAAAAGGACAGGGTTTGGAAGGGCTCCC
AGATAGCTGGACACATGAAGGTTCCTGGAGGGTGGTGCCCCAGAGGGGCA
TGGAAGCTCCACACCCCTTCTCACATGCTTTGCTCTGCGCATCTCTTCAT
CTGGTGTTCATCTGTATCCTTTGTAATATCTTTTAGAATAAACTGGTAAA
CTTAAGTGTTTTCCTGAGTTCTGTGAGCTGCTCTAGCAAATTCACGGAAC
CCGAGGGAAGCAAACCCAGATTTATAGCCATCAGTCAGAAGCATAGGTGA
CAACCTACCACTTGTAACTGGCACCTGAAGTGGGAGGCAGTCTTGTGAGA
CTGAGCCCTCAACCTGTGGGATCTAACGCTAACTCCAGGTAGATAGTGTT
GGAGTGAATTAGGACACCCAACTGGTGTCGGCTGCTGGAGGACTAGTGGT
GGGAGAAATCCCCAAGCATTTCGGTGACTAGAGGTCACAGAAGAACTCAG
TGTTGAGGTGTTGTGACAGTATGGTAGGGAAAACTGCGTCTGGTTTTTTC
CTTTTACAATCAGTTAAATATTTAACACAAGTCTACTGTATATTAGTAAA
AGGGTTACATTTTTTAATGTCTTGACAGTTGCACTTTGACAACTTCCATA
TCAATCACTTTTTTTCGTGTCCGTTTGGAACCAAAATCACTTGGGATACC
ATGAACCAGGCTGCAGCGTATTCCCCAGGCCTTGAAAGCTTGGAGGCCAT
TTTGCCAGCCNTAATCCCTGTGAATACCAGGCTTCGTGGATTTAAAAAAT
AGACTTGAGGCCAGGCCTGGTGGCTCACACCTGTAAGCCCAGCACTTTGG
GAGGCAGAGGCGGATAGATCACAAGGTTAGGAGTTCGAGACCAGCGTGGC
CAACATGGTGAAACCCCGTCTCTACTAAATATACAAAAAAAAATTAGCCG

GGCGTGATGTTACACGCCAGTAGTGCCAGATACTCAGGAGGCTGAGGCAG
GAGAAATACTTGAACCTGGGAGGCAGAGGTTGAAATGAGTCAAGATCGTG
CCACTGCACTCCAGCTTGGGCGACAGAGTGAGACTCAGTTTTCAGGGGAG
TTAAAACAATACAAAAAAAGAAAAAGACTTGAACAATGAGGCTCCACTGG
ATGGATTTAGGGGAATTACAGGAAGCAGGACCTGACGGTGCAATGCCACA
CTCCACCTGTCCAGAATTGGACCTCACCAAGGGAGGTCTGTGGGGACAGG
GAGAGGCCCTCTGCCTCCACCCCCTCCTCTACTCCCCAAACCCTGAGTCA
GGCTGAATGTAGTAAACCTGGAACAGAAAAGTTCAGTTTGGCAATAGGTA
TCTGAAGGACTCCAGGTGCTTCTCCCTTGATTCAAAATTTTACTTATAAA
AAAAATTATAAGAAAATTCTACTTAAAAGAAATAATCAGGGAGGTACAAC
AAATTGTACTTTTTTTTTTTTTTTTTTTTTTTTGAAATGGAGTCTCACTG
TTGCCCATGCTGGAGTACAGTAGTGTGATCTCGGCTCACTGCAACCTCCG
CCTCCTAGGTTCAAGTGATTTTCCTACTTCAGCCTCCCAAGTAGCTGCGA
TTACAGGTGTGTGCCACCACACCCGGCTAATTTTTGTATTTTTGGTAGAG
ACGGGGTTTCACCATGTTAACCAAGATGGTCTCGAACTCCTGACCTCAGG
TGACCCACCTGCCTCAGACTCCCAAAGTGTTGGGATTACAGGGGTGAGCC
ACTAAGCCCAGCCATTGTACATATTTTGTGGGTATTTACTAAAACATTAT
TCAAAATAGTAAAAAAAAAATTGAAATAAACTGGGACTGGTTAAATAATT
TTGGGTACAACCACATGATGGAATACTATACAGCCATTAAAAATTACATT
GAGGCCAGGTGTGGTGGCTCATGCTTGTAATCTTAGCACTTTGGGAGGCC
AAAGTGGGAGGATTGCTTGGACCCAGGAGCTCAAGACCAGCTTGGGCAAT
GTGGCAAAACCCTGTCTCTAAAAAAAAAATACAAAAAAAATTAAAAAGCT
GGGTGTGGAGGCACACACCTCTAGTCCCAGCTACTCAAAGGGCTAAGGTG
GGAAGATCACTTGAACCGGGGAGGTCAAGGCTGCAGTGACCCAAAATCGG
GTCATTGCACTCCAGCCTGGGCAACAAAGCAAGACCCTGTCTCAAAAAAA
AAAAAAATACATTGAAGAATATCTTACGGTATGGATAAATATTCATTTTA
CAGTGATAGATGCAAATAAAAGCAAATTACAAAATATACAGTTTAATTCC
AACTTTGATACTACATATGTATATATGAATACATGCATATGTTATGTATG
TATATGTAAATATAACAATATATGTTCTATATATGGATATTATATATTTA
CACATACATACACACATATATAATATCTTCTCTAGAGAGCAGAAAGAGAG
TAGACAGATAATGAAGATAGGATACAACTCCAGTCCAGCTCAACCTAGGG
GACTTGTTTTAAAGCCTCAGGAGAGAGAAGTTGGGACTAGAAAGCAAGGC
AGCTATTTGTAAGCATCTTTGTGTTTCATGCTATTGGGGTGGGAAACAAC
AGCACAACTTTTGAAAGCCCCTTTCTACTCACCCCACAAACTGCAGAGCA
GCTTTAGGACCCTCAGAGTTCAAGAAGACCATTTGCAGAGTAGAAGAAGT
AAAAACATGTATGAACTTGACCCTGAGCTCATGGACTGTGCCATGAGGGA
AATTCCTAAAACAGCAGGAGAGGCCCTGGAGGAAGGCAGAGGCCCTGCAT
CAGCAAGTCCAGGCAAAAGCCTGCATTCCATAGATGCTCATCTCTCTGGC
TGGTGAGGTCTAAAGACGTTTGGTCTCAATATTAAGTCTCGTGAGAGAGG
TCACAAACCCAGTCCCTTGGCCACAAAAGGAAATAAATTCTGGCTTGAGA
CATTAGGGAGGAACAGGGCAAGGGGAGGTTCAAGAAAGTTTTAATGGATG
AGATGATATTTAAGCAAGGCCCTGGAAAATGAGAATTTCAACCAATAGCC
ATATGGTAGGTCAGAAAGCAAAGATAAGGAGGGGGCAAGTGCAAGGGGCA
ACATCAGATATGACCAGGGTGTCGTGGGGCATGGCTGATGGAGAAGAAGA
TTAGACTGGAGTTTGGGAATGCCACAGTATCGAGGTTGGATTTAATCCTA
TGGGTAATAAAGCCAACTGTTCAACCCCCAACCCACTTGCAATATGGCTC
CAAAATAGCAGGTGTTTGATAAAATGACTACTTTTACTCTACTATTCCCT
CCCTCTTAAGAAGAAAAAGAAAGTGGAGGCTCAGAGAAAGGCAGTGGCTT
GTCCCAATCACACTATGATTTGGCCACAAAACAAGAACGAAATGTTACAC
CCAAAAATGCTGCCTCCACCTCCCTTCCTTGCTTTCCTCCCTGCTGGACT
ACAGACTATCTCAAGAGTGACGTACACCATCAGGGCTTCAGCTTTTCCCC
GAAACAATGCCAAAATATTAGCCATACGTCACTGTAGTAAGAGCCCTGAA
TTGGGAATCCCAGCTTTGACGCAGACATGCTGATTGACTCTGTGACCATT
CTCTTCACTTCTCCACTCTATTCTTCCCCACCTGTAAAGTGAGGTCCTTT
CCAGTTATAAAAACAGATGATGCTATTGTCCTGTTTTGTATCTAATCTTG
CTGTGTTATAAAAAAAAAATAAGGCTCTGTACATTCATCTTGGCCAATTC
CCTTCTTATCTCTACTTCCCACAGCCCCTTTTTCTACAGAAAACCAGCAT
TGTTCTTCTGGATCCATCTCTTAAGAAAGCGCTTTGCCTCCCCGGTTATT
TAGGTGATAAGAAGTGTCCTAGATGACAGCCCTGGAATGGGCTGGAGGCA
ACAAAAAAGCAAGTGAAATAGACAGTTACAGCGACGACAATAATAACAAC

FIG. 3 (22 of 52)

CAACACCTCTCACTAAAGAGAAAGAAATAAAAAAGAAAATTAAAATCTGC
CGCAATGCCCACACAGTCATTGAATAACTGCATGTGTACAGCACTTGGTT
ACTTTTACATACTTCATATTTTAGCCTTCATAGCAGCTCACAGGGGTGGA
TTTAATTTTTAGTCCAACTCCTGTCACGGTGCCTGGCACAAGTATAATAA
ATGTTCTGTGAATAAATGACCCTCTTTTTAGATGAGGAAATCGAGGCTCA
AGGAGAACAAGCAATGTAATGTCCCCCTCCTGTTCAGCCATCTGCCTTTC
ACGCCACTGAATGCAGTAGTCCTCAGTGCCCTGAACTTGACCCTCTTCTG
CTTTTCGGACTGGTCCTTCTAATCCCGTTGTGACTCACTACACCACCTCT
CCTGCATATGACATCTACATTTTAAAACAAACCGTATGGAAATAACACAT
TAGTCGGCTTGTTCCCCCACCCCCGCAAAAAAAAAGGCCTCTTTATAACA
GAAACTTCTCAGGCTGGTAGGGGAATTTTATTCCCCCATTTATGGTAGAA
AGGCCCTAACCTTGGACCTCACGCCATAGCTATTCACATGGGGGAATGAT
GAATAACATGGGGAGCAGCATGTAAATATCATTGAGCCGTAGTCCAGACC
TATAACACATC

>Contig31

GGGGGAGCTGCATGTGCCTGTCGAGATCTGGGGGAGGAACAGGAAGATCA
AGAGTTCTGTGTAGGACATGTTAAGTTGAAGGTGCTTACAGGATAGCCAG
ATGAAGCATCAGGTGTGCAGTCAAAGATATGAGTCTGGAGCAGCACATCC
TAAGTCACCTCCTGCACCAACACAGAACTTCCAGGCCACTCACTTGAGCT
CTCCCAAATAGTTTCCAAGTGTCATTATGTTAATAACCTATGAGCTTGAA
CACCAGATTCAAACCCCACTGCATGGCTTTTAAAGACCATCTCAAGGGCT
TGACACTCCAGGGAGCCAACTAAAGATGCCTGGTCCTACCATCAACCTCC
ACCCCATTTTTTATAGAAAATGTTTCTACCTGTCCTAAGGCAGGGTCCTG
CCCCACTCCCAGGCCCCTTTAGATCCCCAATATTCCTCCTCCCTGAACCA
AAACCCTCATCATCTTCCAGCATGGGTGGGGCCTCCATTCTTGCTTCTGC
TCCCCTGAGCAGAAGCAAGTTTCTCCCAACTTGACCTGATTCTCCTCCTA
AGTACCAGTCACTGCTTTGTTTCTGGAATGAGAGAAAAAGACAGAGTGAG
AGAGACAATCCAGAACTCTTGCTCACTCACAGCTAGGCTGGGCATCTGGG
AGGATGGCTGTGTCCATGGGAACCTGGGAAAAGCCACACCCTTGGCACCC
TGGTCACCCACCTGTCTCCCTGGCAGATTCCGCACTGCTCTCTTGCACCC
TCTACCAGGGCTAACCGGCCTGCTCACTCTCCCCAGCATGTCTTCCCACG
CCCACTCTCTAATTATTACATTCCCTTCACATAAACTGCCCTTCTCTCCC
AATCACCACATGTTCACTTCCCACCCAGCTGTCAAAGTCTGGCTCAACCT
CATTCTTGAAAAGGAAAAAACAAACAAACAAACAAACAAACAAGCAAAAA
ACCTATGATGGATTAAGAACACACTTCATTCCAGGAACATGCTTATCTCC
TCTAACTCTCACAACAACTACAGCAGGTAGGTGTTATCACACCCATCTCT
CAGGTGAGAAAACAGGCTCAACGAGTGCAGGAGGACACAGCAAGTCAGTG
ACAAAGCTTAAATTCAAGCCCAAGCCTGTTGGCAACCAACGTCTGTACCC
TTGATAGCTACCTCATTTACCACCAAATCCAGTGGCCTCAGGCCTGGCTG
CACACTGGGATCACCTGGTGCCCAGACCACATCTTAGACCAGTCATACAG
AATCTCTTGGGCTGGGATCCTCCACGGTACATTTTAAGGGTCCCCAGGTG
AGTTCCACCATGGACCCAGAATTGAGGACCCAATACCGTATACCATCTCC
TTCTTCATCTCTTCTAAGGCATCTCTTACTCGCTGTGCACTCCCATACCA
CTTTGTTCAATCATCCAATCATTCATTCATTGAGTCAGTTAGTCAGGAGC
TACTCACTAGTCCCCTGCCAGGTCCTAGTCATGACATAGGGCTCTGGGGA
CCAACAAGAAGCAGGACCCATGCCTCCTGCTCTCATGGAGCTTGCTCTGC
AGCAGAGGAAGCAGTCAGTGAGATGTAGCAAATGTGAAATGTGCACAGAT
GGGAAAAGCAAAACTTTAAAACTTTTAGGACAAAATACACAAGAAATCTT
TGCAACTTTGGGACAGGAAGGAACAACATTCCTTACACATGACACCAAAG
GAATCAACCATAAATAAAAAGGTGATCAATTTGACCTCATTTAAGTGTTA
AGCTTTTTTCATTGAGAGACACCATTAAAAATTAAAAATACATGCCACAA
ACTGGGATACAATATTTACAACACTTATGTCTCACAAAGGATTAGTTTTC
AGAATATATAAAGAACTCCCGGCCGGGTATGGCCGCGCACGCTGGAATCT
CAGCACTTTGGGAGGCCAGCGGATCACATGAGGTCAGGAGTTCAAGACCA
GCCTGGCCAACATGGCAAAACTCCGTCTCTACTAAAAATACAAAAATTAG
CCAGGCATGGTGGCGGGCGCCTGTAATCCCAGCTACTCAGGAAACTGAGG
CAGGAGAATCACTTGAGCCCAGAAAACAGAAGTTGCAGTGAGCTGAGCTC
ACATCACTGTAAGCCTCGGTGACAGAGTAAGACTGTCAAAAAAACGAAAA
CAAAAACAAAAACTCCTACAAATAAATAAGAAAAAAATAGCCCAGCAGGA
AAAAGTATATACATTTCATAAAAGAATAAATACATTCTGTCAGTTTTCTA

FIG. 3 (23 of 52)

ACATATATTTTTTAAGAGTAAATACAAATGGTTAGGAAACATTTTTTAAA
ATGCCCAACCTCATTAAAAATTATAGAAGTGAAAATTAAGCCACAATAAG
ATACGATTTTATACCAAATACAGTGTCAACACTTTGCAAGTCTGACCTCA
CCAAGTGTTACCAGACGTGTGCACTGACGTGGCTGCTGAGATACTGATGG
TGGGTTTGTAAATCTGTACTACAAACAATTGCAATAAAATGTAATAAATA
TACAATAGGTGGAGCAGGAAGTGACCTGCAACCATATAGCAGATAGGGCA
GGAAAAAGCCTATGAAAGCTGACATCAAAGGGATAAGTTCCAGTTACCCA
GCTGAAGGGAAGGAGGGTGTTTCAGATAGAGGAAGGATAAGCATGACCTA
TTCAAGGCCAGTGAAAGAAGCGTGCAACGGCCAAGTCAGGAGAACCTGAA
ATTGTGTCAAAGAGCTTGGATGCAAAGAGCCGTGGGAGACTATTGGGGGT
TTTAAGCAGGGATATAATATTCATTCAAGCATGCAGTAAAAGGTCACTGG
CACCTGCCATGGGCCAGGACTCGGGCTCTACATGATTGCGTCTGTTTTGG
AAATATCACCCTGGCTGTGAGATGAAGAACAGGTAGGAGGGTCACAAAAC
TTGAAGCAGAGAGACTGTTGAGGAAGTAAGCTGTTTTTGTGTGGACTGTG
GCAATCACAGAGGCAGAGGATATAAATGCACAGAGACACAAGGCATGTGG
GAGGCAGAAGGAATCAAATACAATGAGTGATCAGATGTGGGGTTAGAGTG
GTGAGTGAGAAGACATACTCAAGGTGACACGCCCAGGTATCTGGGTGGAT
GGTAAGACATTCATGGACTAGGATCGAGGAANGAGGTGGGGAATGGGACC
ATACCTGCAGTTTATAAGGGGTGGACGAGGGAAGATTATGCGGGAGACTG
AGAGAGGAATAGACAAAGGAATCCCGGTGCAGTATTACAGAAACTGGGGT
GGGAGGGGGTTGTANTTCAAAAAGGAAAGAAAATTGTCAAATAGTATGAA
ATGCTGCAGAGAAACTCACGGATTTTTTTTTTAAGCTTAGAATTATTCAT
TGACTATGTGAATAAGAATAACTTTTATGAAAGAAGTTTTGCTTAAGTAG
TAGGAAGAAGCAAAATTGTTGAGGGCTGATGAGTGGGAGGAGAAGTAATT
GAAGGCACTCTTTCAAGAGAAACAAAGCAGAAGGTGAGGAGAATACTAAT
GAAGGAGTTACGGCCTTCACTATTTTGTTTTGCTTTAGATAAGCAAGACT
TGAGTGGGTCTGGTGAGGAGAAACAAGTAGAGTACAAAGTTAAAGGAGAG
ACAGACAGAGATAGAGATAGGGACAGAGAGAGAGACAGAGACAGAGCACA
AAAGAGCAAGGTCCCTGAGAACACGGGCCTTCTGTTTAAACCCCAGCCAG
ATGTATTGCAATTCAATTCCAGTACTAACCACCCAGAGTTTGTGTAGACT
CTACAAGTTAAAGAGCATGGTCCCCAACAAGACTGCTTCTACGTCAGATG
CCAGGCACACTTCAGGGGTCCCCAAGCCACTCATGTTTTTTGAATGACTG
CCATAAGTTCAAAAATTCCCACAATTCTCTCAGATTCAATAACTGGGTAT
AACCACTCATAGAACTCAAGAAAATGCTATCATTATTATTACAATTTTAT
TATAAAGGATACAAATCAGAAGGACTAGCCAAATGAGGAGACACATAGAG
AGAGGACTAGTAAAAAACAGAGCTTCTGCGTCCTACCTTCAAGGAATCAG
GATGCACCACCCTCCCAGCACATCAAGTGCTCATCAACCAGGAAGTTCCT
CTGAGCTCCAATGTCCAGAGATTTTAGGGAGGATTCATTACATAGGTATC
ATTGATTAAATCATTGGCCATGTACTTGAACTCAATCTCCAGTGTCCCTC
TTCTCCCTAGAGGTCTGAAGGGTTGGCTAATATCATGTGGCTCAAAGCCC
CAACTCTAATTACCTTTTTGGTCTTTTCAGGGACTAGACCCCATCCTGAA
GCTATCTACAGGCCCTGCCATGAGTTAGCTCATTAACATAACAAAGACAC
TTATATTACTCAGAAAATTCCAACAGTTTTAGAAGCTCCATGTCAGGAAC
CTGGGACATAGATCAAATTCTTTTTTTTTTTTTTTTTTTGGAGACAGGGT
CTTGCTGTGTTGCCCAGGCTAGAGTGCAAGGACAGATCACAGCTCAATGC
AGCTTCAACTTCCCAGGCTTAAGTGACCTTTCCACCTTAACCTTCCAAGT
ATCTGGGACCACAGAAAATGGCTAATTATCCTGGCTGATTTTTAAACTTT
TTTTTTTTGTAGGGATGGGATCGCCCTGTGTTGCCAAGGTTGGTCTCAAA
CTCCTGGGTTCAAGCAATCATTCTGCCCTGGCCTCTGTGATGGTTAATAC
TGAGTGTCAACTTGATTGGATTGAAGGATACAAAGTATTATTTTTGGGTG
TGTCTGTGAGGGTGTTGCCAAAGGAGATTACATTTGAGTCAGTGGACTGG
GAAAGTCCACCCTTTCCCAGTGGACTGGGAGACCCACCCTCAATCCAGGT
AAACACAATCTAATCAGCTGCCAGTGTGGTCAGAATAAAAGGAGGCAGAA
GAACAGGGAAACACTAGACTGGCTTAGTCTTCCAGCCTACATCTTTCTCT
CATGCTGAATGCTTCCTACCCTCGAACATCAGCCTCCAAGTTCTTCAGTT
TTTGGACTCTTGGACCTTCAACCACAGATTGAAGACTGCAGTGTTGGCTT
CCCTGTTTTTGAGGTTTTGGGACTCAGACTGGCTTCCTTGCTCCTCAGCT
TGCAGATGGCCAATTGTGGGACTTTAACTTGTGATCATGTGAGTCAATAT
TCCTTAATAAACTCAGATATATATATATGTATCAGACATATATATATATC
CTATTGTATATTATATACAGATATATAATATCCTATTATATACAGATATA

FIG. 3 (24 of 52)

TAATATCCTATTATATACAGGTATATATATATATATGTATCATATATATA
TATCCTATTGGTTCTATCCCTCTTGAGAATCCTGACTAATACAGCCTCCC
AAAATGCTGAGATTACAGGAGTGAGCCACAGCCACCATGCCCAGCCCCAA
ATTCTTAATTATACAACAATGGGTCCAGAGATCAGGGCCTGGGTAGGATG
CAGCAATAAGAAAACAGATGGTGGATGGGGACACATGTTGGAAGTGTGGC
AGGACATGGCTGAGGGAACTCATAGGATGGTGTCTATTTTCATGGCTGAG
TGTGAGGAACAGCATAAGGTCAAAATTTCAGGTCAATGGTGAGTTTTTTA
AATTGTTGCTGTGAACCCCAAAAAATCTGACCCAGGTCTCAGTTAATTTAG
AAAGTCTATTTTTCCAAGGTTGAGAACACCCACCCACTCACGACAAGAGC
ATCAGGAGGTCCTGACCACATGTGCCCAAGGTGGTAAGAGCACAGCTTGG
TTTTATATATTTTAGGGAGACGTAAGTCATCAATCAATATATGTAAGATG
TACACTGGTTCTGCCTAGAAAGGCAGGACAACTTGAAGCAGGGAGGGGGC
TTCCATGTCACAGGTAGGTGAGAGACAAACAGTTGCATTCTTTGAGTTTC
TGATTATCCTTTCCAAAGGAGGCAATCAGATGTGCAATTATCTCAGTGAG
CAGAGGGATGACTTTGAATAGAAAGACAGGCAGGTTTGCCCTAAGAAGTT
CCCAGCTTGACTTTTTCCTTTAGCTTTGTGATTTGGAGGCGCCAAGATTT
ATTTTCCTTTCACATTTCCCCCCCTTTCTTTTTAAGAATCTTTTAAAGAA
AGCTTTTAAAAAGAAAATGAGTCTCTGGTCCCAGGTTTCATCTGAATTCT
CGAGGGGAGGATGGTTTATCCTAAACGGGTGGTTCTGAATTTTGAGAAAG
TGCATTGTAC

>Contig32

AAAAGCCATACGAATGAGGAAGAATTAAGGGCCAGAACAAAACAAGAAGA
TGAGGGAAAGTTTGGAACTTCTTAGAGACTGGCTAAATGGTTGTGACCAA
AATGCTGATAGTGATACGGACAATGAAGTCCAGGGTGACAAAGTCTCAGA
TGGAAATGGGGAATTTGTTGGGAACTGGGCAAAGGTCACCCTTGCTATGA
CTCAGCAAAGAAATTGGGTGCATTGTGTTCATGTCCTGGGGATCTGTGGA
AGTTTGAATGTAAGAGTGATGACTTACGGTAGGGTATCTAGTGGAAGAAA
CCTCTAAGCAACAAAGTGTGTTGCTTAGAAATTTCTTTCTTTCTTTTTTT
TTTTTTTTTGAGCTGGAGTTTTGCTGTGTCGCCCAGGCTGGAGCGCAGTG
GCGCAATCTTGGCTCACTTCAAGCTCTGTCTCCTGGGTTCATGCCATTCT
CCTGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCTGCCACCATAC
CTGGCTAATTTTTTAGTATTTTAGTAGAGACGAGGTTTCACCATGTTAGC
CAAGATGGTCTCAATCTTCTGACCTCGTGATCCACCCGCCTTGGCCTCCC
AAAATGCTGGGGTTACAAGCATGAGCCACCCCGCCTGGCCTGCTTAGAAA
TTTCTAAGCCAGGATATGGCCTGTCTGCTTCTAACAGCCTGTGCTCAGGG
GTAAAGAAATGACTTAAAGTTGGAACCTATGTTTAAAATGGAAGTAGAGT
CTAAAAATTTGGAAAATTTGCAGCCTGGCCTTGTGGCAGAGAAAGAATCC
AAGTAGGCTGCAGAGCAATCATTGCTAGAGAGATTAGCATGACTAAAAGG
GAGCCAAGTGCTAATATTCAAGACAATGTTAAAAAGGCCTTGAGGGCATT
TCAGAGATCTATGAAGCAGCCCCTCCCATCACAGGTGCAGAGGTTTGGTG
CACTAGGCCCAGAGGTTTTATGGGCCANNGCCAGGGCCACACTGCTATGC
ACAGCTTTGGGACACTGCTGCCCGCATCCAGGCCACTCTGCTCTGGCTCC
ACCCTTGGCTCAAACGGGCCAAGATAGAGCTTGGACCACTGCTCCCGAGG
GCACAAGCCATAAGCCTTGGTGGTTTCCATGTGGTGTTAAGCCTGCAGGT
GCCCAGAATGCAAGATTGAGGGAGCTTGGGCACTTCCACCTAAATTTCAG
AGGATGTGTCAGAAACCCTAGGTTCCCAGGCAGAAGCATGATACAGGGGC
AGAGCCCTTGCAGAGAACCTCTACTAGGGCAATGCCAAAGGAAAATGTGG
GGTTGGAGTCCTCACACATGGTCCCCACTGGGGCACTACCTGGTGATACT
GTGGGAATGGGGCTGCTGCCCTCCAGACCCCAGAATGGTAGATGCACTGG
CAGCTGGCACCCTGAGCCTGGAAAAGCTGCAGGCACTCAACTCCAACCCA
TGAGATCAGCCACATGGGCTACTCCCAGGGAAGCCCACAGAGGCAGGGCT
GTCTAAGGCCTTGGGAGCCTACCCCTTGAACCAGCTTGCAGGACATGGAA
TCAAAGATTATGTTGCAGCTTTAAGGCTTAATGTTTTCCCTGTCAATTTC
AGGCTTGTGTGGGACCTGTTGCTTTTTTTTTTTTTTTTTTTTTTTTTGGT
CACAGGTGTTTGAACCAGAACAATTCCATCTTGAATAGGGGCTGGGTAAA
ATAAGGCTGAGACCTACTGAGCTGCATTCCTAGGAGGTTAGGAATTCTAA
GTCACAGGAGGAGATAGGAGGTCGGCACAAGATACAGGTAGCGAAGACCT
CGCTGATAAAATAAGTTGCAGTAAAGAAGCCAGCCAAAACTCACAAAGCC
AAAATGGTGATATGGTTTGGCTCTATGTCCCCACCCAAATCTCATCTCAA
ATTATAATTCCCATAATCCCCACATGTTGAGGGGAGGACCTGGTTGGAGG

FIG. 3 (25 of 52)

TGATTGGATTATGGAGGCAATTTCCCCCATGCTGTTCTGGTGATACTGAG
TGAGTTCTCATAAGATCTAATGGTTTTATAAGTGTTTGGAAGTTCCTCCT
ACACACATGCTCACACTCTCTCCTGCAGCTTTATGAAGAAGGTACTTGCT
TTCCTTTCTGCCATGATTGTAAGTTTCCTGAGGCTTCCCAGCTATGCAGA
ACTGTGAGTCAATTAAACCCGTTTTCTTTATACATTACCAGTCTTGGGCA
GTTCTTTACAGCAGTGTGAGAACTGCTGGCGATGAGAGTGACCTCTGGTT
GTCCTCACTGCTCATTATATGCTAATTATAATGTATTAGCATGCCAAAAG
ACACTCCCACCATGACCCCAACAGTCATGCCTGTGCCGGTCTCAGCACCA
TGACAGTTTACAGATGGCATAGCAACGTCTAAAAGGTACCCCATATGGAC
TAACAAGGGGAGGAACCCTCAGCTCTGGGAAGTGCCTACCTCGTTCCCAG
AAAGCTTGTGAATAATCCACTGCTTGTTTAACATATAATTAAGAAATAAC
TATTAAGCATCCTTAGTTCAGCAGCCCAAGCTGCTGTTCTGCCTATGGAG
TAGCCATTCTTTATTCCGTTACTTTCTTAATAAAATTGCTTTTACTTTAC
TGTATGTACTCGCCTGGAATTCTTTCTTGTACGAGGTCCAGAGCCCTCTC
TTGGGTCTGGATCGGGACCCCTTTCTGGTAACATTTTGACCAATTTCTCC
CTTCTGGAATGGGAATGTTTACACAATGACTGTATCACTTTTGAATCTTG
GAAGTAAATAATTTGTTTTTGACTTTACAGCCTCATAGGTGGAAGGAACT
TGACTTGAATTTCAGATGAGACTTTGGACTTTGGGACTTTTGGGTTGGGG
CTGGAATGAGTTAAAAGTTGGGGGGATTATTGGGAAGGCACGATTTTATT
TTGCAATATGAGAAGCACATGAGATTTGGGGGACCAAGGGTGGAATAATA
TGGTTTGGATGTTTGCCCCCTCCAAATCTCACATTGAAATGTAATCCCCA
GTGTTGAAGTGAGGCCTGCTGGAAAATGTTTGGATTACAAGGCTGTCGAG
CACATTGGATAAGACGTGTAGGNCCC
>Contig33
CGCAGCTCGCTGGTTAATTCTGTGGCTCCTGTGACCACTATTATAGCACC
AGGTCTATGACCAGGAGAATTAGACTGGCATTAAATCAGAATAAGAGATT
TTGCACCTGCAATAGACCTTATGACACCTAACCAACCCCATTATTTACAA
TTAAACAGCAACAGAGGGAATACTTTATCCAACTCACACAAGCTGCTTTC
CTCCCAGATCCATGCTTTTTTGCGTTTATTATTTTTTAGAGATGGGGGCT
TCACTATGTTGCCCACACTGGACTAAAACTCTGGGCCTCAAGTGATTGTC
CTGCCTCAGCCTCCTGAATAGCTGGGACTACAGGGGCATGCCATCACACC
TAGTTCATTTCCTCTATTTAAAATATACATGGCTTAAACTCCAACTGGGA
ACCCAAAACATTCATTTGCTAAGAGTCTGGTGTTCTACCACCTGAACTAG
GCTGGCCACAGGAATTATAAAAGCTGAGAAATTCTTTAATAATAGTAACC
AGGCAACACCATTGAAGGCTCATATGTAAAAATCCATGCCTTCCTTTCTC
CCAATCTCCATTCCCAAACTTAGCCACTGGCTTCTGGCTGAGGCCTTACG
CATACCTCCCGGGGCTTGCACACACCTTCTTCTACAGAAGACACACCTTG
GGCATATCCTACAGAAGACCAGGCTTCTCTCTGGTCCTTGGTAGAGGGCT
ACTTTACTGTAACAGGGCCAGGGTGGAGAATTCTCTCCTGAAGCTCCATC
CCCTCTATAGGAAATGTGTTGACAATATTCAGAAGAGTAGGAGGATCAAG
ACTTCTTTGTGCTCAAATACCACTGTTCTCTTCTCTACCCTGCCCTAACC
AGGAGCTTGTCACCCCAAACTCTGAGGTGATTTATGCCTTAATCAAGCAA
ACTTCCCTCTTCAGAAAAGATGGCTCATTTTCCCTCAAAAGTTGCCAGGA
GCTGCCAAGTATTCTGCCAATTCACCCTGGAGCACAATCAACAAATTCAG
CCAGAACACAACTACAGCTACTATTAGAACTATTATTATTAATAAATTCC
TCTCCAAATCTAGCCCCTTGACTTCGGATTTCACGATTTCTCCCTTCCTC
CTAGAAACTTGATAAGTTTCCCGCGCTTCCCTTTTTCTAAGACTACATGT
TTGTCATCTTATAAAGCAAAGGGGTGAATAAATGAACCAAATCAATAACT
TCTGGAATATCTGCAAACAACAATAATATCAGCTATGCCATCTTTCACTA
TTTTAGCCAGTATCGAGTTGAATGAACATAGAAAAATACAAAACTGAATT
CTTCCCTGTAAATTCCCCGTTTTGACGACGCACTTGTAGCCACGTAGCCA
CGCCTACTTAAGACAATTACAAAAGGCGAAGAAGACTGACTCAGGCTTAA
GCTGCCAGCCAGAGAGGGAGTCATTTCATTGGCGTTTGAGTCAGCAAAGG
TATTGTCCTCACATCTCTGGCTATTAAAGTATTTTCTGTTGTTGTTTTTC
TCTTTGGCTGTTTTCTCTCACATTGCCTTCTCTAAAGCTACAGCCTCTCC
TTTCTTTTCTTGTCCCTCCCTGGTTTGGTATGTGACCTAGAATTACAGTC
AGATTTCAGAAAATGATTCTCTCATTTTGCTGATAAGGACTGATTCGTTT
TACTGAGGGACGGCAGAACTAGTTTCCTATGAGGGCATGGGTGAATACAA
CTGAGGCTTCTCATGGGAGGGAATCTCTACTATCCAAAATTATTAGGAGA
AAATTGAAAATTTCCAACTCTGTCTCTCTCTTACCTCTGTGTAAGGCAAA

FIG. 3 (26 of 52)

TACCTTATTCTTGTGGTGTTTTTGTAACCTCTTCAAACTTTCATTGATTG
AATGCCTGTTCTGGCAATACATTAGGTTGGGCACATAAGGAATACCAACA
TAAATAAAACATTCTAAAAGAAGTTTACGATCTAATAAAGGAGACAGGTA
CATAGCAAACTAATTCAAAGGAGCTAGAAGATGGAGAAAATGCTGAATGT
GGACTAAGTCATTCAACAAAGTTTTCAGGAAGCACAAAGAGGAGGGGCTC
CCCTCACAGATATCTGGATTAGAGGCTGGCTGAGCTGATGGTGGCTGGTG
TTCTCTGTTGCAAAAGTCAAGATGGCCAAAGTTCCAGACATGTTTGAAGA
CCTGAAGAACTGTTACAGGTAAGGAATAAGATTTATCTCTTGTGATTTAA
TGAGGGTTTCAAGGCTCACCAAAATCCAGCTAGGCATAACAGTGGCCAGC
ATGGGGGCAGGCCGGCAGAGGTTGTAAAGATGTGTACTAGTCCTGAAGTC
AGAGCAGGTTCAGAGAAGACCCAGAAAAACTAAGCATTCAGCATGTTAAA
CTGAGATTACATTGGCAGGGAGACCGCCATTTTAGAAAAATTATTTTTGA
GGTCTGCTGAGCCCTACATGAATATCAGCATCAACTTAGACACAGCCTCT
GTTGAGATCACATGCCCTGATATAAGAATGGGTTTTACTGGTCCATTCTC
AGGAAAACTTGATCTCATTCAGGAACAGGAAATGGCTCCACAGCAAGCTG
GGCATGTGAACTCACATATGCAGGCAAATCTCACTCAGATGTAGAAGAAA
GGTAAATGAACACAAAGATAAAATTACGGAACATATTAAACTAACATGAT
GTTTCCATTATCTGTAGTAAATACTAACACAAACTAGGCTGTCAAAATTT
TGCCTGGATATTTTACTAAGTATAAATTATGAAATCTGTTTTAGTGAATA
CATGAAAGTAATGTGTAACATATAATCTATTTGGTTAAAATAAAAAGGAA
GTGCTTCAAAACCTTTCTTTTCTCTAAAGGAGCTTAACATTCTTCCCTGA
ACTTCAATTAAAGCTCTTCAATTTGTTAGCCAAGTCCAATTTTTACAGAT
AAAGCACAGGTAAAGCTCAAAGCCTGTCTTGATGACTACTAATTCCAGAT
TAGTAAGATATGAATTACTCTACCTATGTGTATGTGTAGAAGTCCTTAAA
TTTCAAAGATGACAGTAATGGCCATGTGTATGTGTGTGACCCACAACTAT
CATGGTCATTAAAGTACATTGGCCAGAGACCACACTGAAATAACAACAAT
TACATTCTCATCATCTTATTTTGACAGTGAAAATGAAGAAGACAGTTCCT
CCATTGATCATCTGTCTCTGAATCAGGTAAGCAAATGACTGTAATTCTCA
TGGGACTGCTATTCTTACACAGTGGTTTCTTCATCCAAAGAGAACAGCAA
TGACTTGAATCTTAAATACTTTTGTTTTACCCTCACTAGAGGTCCAGAGA
CCTGTCTTTCATTATAAGTGAGACCAGCTGCCTCTCTAAACTAATAGTTG
ATGTGCATTGGCTTCTCCCAGAACAGAGCAGAACTATCCCAAATCCCTGA
GAACTGGAGTCTCCTGGGGCAGGCTTCATCAGGATGTTAGTTATGCCATC
CTGAGAAAGGCCCCGCAGGCCGCTTCACCAGGTGTCTGTCTCCTAATGTG
ATGTGTTGTGGTTGTCTTCTCTGACACCAGCATCAGAGGTTAGAGAAAGT
CTCCAAACATGAAGCTGAGAGAGAGGAAGCAAGCCAGTTGAAAGTGAGAA
GTCTACAGCCACTCATCAATCTGTGTTATTGTGTTTGGAGACCACAAATA
GACACTATAAGTACTGCCTAGTATGTCTTCAGTACTGGCTTTAAAAGCTG
TCCCCAAAGGAGTATTTCTAAAATATTTTGAGCATTGTTAAGCAGATTTT
TAACCTCCTGAGAGGGAACTAATTGGAAAGCTACCACTCACTACAATCAT
TGTTAACCTATTTAGTTACAACATCTCATTTTTGAGCATGCAAATAAATG
AAAAATCTTCCTAAAAAAATCATCTTTTTATCCTGGAAGGAGGAAGGAAG
GTGAGACAAAAGGGAGAGAGGGAGGGAAGCCTAATGAAACACCAGTTACC
TAAGACCAGAATGGAGATCTTCCTCACTACCTCTGTTGAATACAGCACCT
ACTGAAAGAACTTTCATTCCCTGACCATGAACAGCCTCTCAGCTTCTGTT
TTCCTTCCTCACAGAAATCCTTCTATCATGTAAGNTATGGCCCACTCCAT
GAAGGCTGCATGGATCAATCTGTGTCTCTGAGTATCTCTGAAACCTCTAA
AACATCCAAGCTTACCTTCAAGGAGAGCATGGTGGTAGTAGCAACCAACG
GGAAGGTTCTGAAGAAGAGACGGTTGAGTTTAAGCCAATCCATCACTGAT
GATGACCTGGAGGCCATCGCCAATGACTCAGAGGAAGGTAAGGGGTCAAG
CACAATAATATCTTTCTTTTACAGTTTTAAGCAAGTAGGGACAGTAGAAT
TTAGGGGAAAATTAAACGTGGAGTCAGAATAACAAGAAGACAACCAAGCA
TTAGTCTGGTAACTATACAGAGGAAAATTAATTTTTATCCTTCTCCAGGA
GGGAGAAATGAGCAGTGGCCTGAATCGAGAATACTTGCTCACAGCCATTA
TTTCTTAGCCATATTGTAAAGGTCGTGTGACTTTTAGCCTTTCAGGAGAA
AGCAGTAATAAGACCACTTACGAGCTATGTTCCTCTCATACTAACTATGC
CTCCTTGGTCATGTTACATAATCTTTTCGTGATTCAGTTTCCTCTACTGT
AAAATGGAGATAATCAGAATCCCCCACTCATTGGATTGTTGTAAAGATTA
AGAGTCTCAGGCTTTACAGACTGAGCTAGCTGGGCCCTCCTGACTGTTAT
AAAGATTAAATGAGTCAACATCCCCTAACTTCTGGACTAGAATAATGTCT

FIG. 3 (27 of 52)

GGTACAAAGTAAGCACCCAATAAATGTTAGCTATTACTATCATTATTATT
ATTATTTTATTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGC
TGGAGTGCAGTGGCGCAATCTTGGCTCACTGCAAGCTCTGCCTCCTGGGT
TCACGCCATTTTCCTGCCTCAGCCTCCCGAGTAGCTGGGACAACAGGCAT
GTGCCACCATGCCCAGCTAATTTTTTTGTATTTTTAGTAGAGATGGGGTT
TCACTGTGTTAGCCAGGATGGTCTCTATTTCCTGATCTCATGATCCGCCT
GCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCG
GCCTATTATTATTATTATTACTACTACTACTACCTATATGAATACTACCA
GCAATACTAATTTATTAATGACTGGATTATGTCTAAACCTCACAAGAATC
CTACCTTCTCATTTTACATAAAAGGAAACTAAGCTCATTGAGATAGGTAA
ACTGCCCAATGGCATACATCTGTAAGTGGGAGAGCCTCAAATCTAATTCA
GTTCTACCTGAGTAAAAAAATCATGGTTTCTCCTCCATCCCTTTACTGTA
CAAGCCTCCACATGAACTATAAACCCAATATTCCTGTTTTTAAGATAATA
CCTAAGCAATAACGCATGTTCACCTAGAAGGTTTTAAAATGTAACACAAT
ATAAGAAAATAAAAATCACTCATATCGTCAGTGAGAGTTTACTACTGCCA
GCACTATGGTATGTTTCCTTAAAATCTTTGCTATACACATACCTACATGT
GAACAAATATGTCTAACATCAAGACCACACTATTTACAACTTTATATCCA
GCTTTTCTGACTTAGCAATGTATTGATGACATTATGCATGCTTAGACCTC
C

>Contig34

GTATTCTATTCTCGGTTATAACACAATCACAGTGATTTGTCATATCTTTC
CAGGATTTGTTAATTTCACTTCTTCAGCTGTTTCCCCCTTGTTGGCTGGA
ACTGATTTTCTATCTTCTGGGAGAATCTTCAGCAAGCCAACTCAGGATTT
GTTGGGTGCATTTTGTCAAGTCTAGGACCCAGGCTCTGGGTGACTGATTT
CCTCTAATTACCGAGCAATGTAAAATGAGGAAGTCTGATTGTGTAAAGGT
GTTAAACTTTTGTGTGACGGCAAAACTTTAATACCATGAATAGAGATTCC
AGAATTTTCCAACTTCTAACGGGATTCCTTTCACTCCCTGACATTAGAAT
GTTAGAAAATCTACCACAAAACATCTGTGAGGCTATCCTACAAGGCCCGT
TTTTCAAAATAGGTTTTTACAAGGATTGCTATTTGGGATGATAGTTTCAG
AAAGGCGCTATCAAAGTTAATTGATGATGTGTGCAAGCTGAAAGTTATAT
GTTAGAACTAGCAGTGATTTCAAAAATATCCCTTTTAGGCTTTTTGCTAA
TATATCTGCTCATTTTCAAAGTTCCCAATATTATAAAACTTTTTAAAGCA
GAAAGAAGAACCCTCCATTTCTGCTGGCCCCTTCCCTGTTCAACTAAAAA
GTATTTTCCCAGGCAATGCTATCCCAGGACTCACACTCCATCCATCCATC
ACCTACCATAAGTTCTTTGAAGGGCTCATTCTGAGCGCTTCCTGAGTGCC
TGGGATCTGTTATTTCTCTCCATTTCTGCTGCTGCATGGTAGTCCAAGTC
CTCCTCCCTTTTCCCCTAGGCCATTTGAATCATCTGCTAATTGGTTTTCC
TGATTGCCACGGAAACTTCCTCCATCCCTTCCTCACATATCAGCCACAGA
AGTATCTCCAAAAAGCAAATCTGGTGACATGAAGCCCTTGCACAAAACCC
ATTCATTACTGGTTCCACACCTCCTTTGTGGATAAGTTCAAGCTCCTGAG
TGTGGCAAGCAGGGCCCACCTGGAATCCCCTGCCCTCCTCTCCTATCCCA
CGCATCAATCTTTCCTGTCTATTTGCAGTTCCTTGAATGTGATATTCTTT
CTAGTCTCTGTGCTTTTGCATAACCTGTTCTTCCTGACTGGAAACTCCTT
CTCCTCCTTGTAGTTTGGCTAATTTCTAGTCTTTCAAGACTCAGCTCATG
CTTCACCCCCTCTATAACAAGTCCTTTCCCAAGCTGGGTGGTGGATGCTC
CTCTGTGCTGTGTGAGTCTTGAACATCCTCAGCAAACCTCAGCTTTGTTT
GCTTGTCTCCCTTGCTGTCAATGCACCTGATTCAGGGCTGGCATATACTG
TTCACCTCCATGACTGGCTCATGGTGGTGCTCCGTGAATATCATCCACCC
AAACGGATGAGAGCTACCATGCCATCACTTGTGACTTCCATCTGGAGCTA
ACCTCCCCCGACAGGAAAGCGTTTCCTTAGGAAAGAATATCTTTGGGTTA
AATAGAAGTAGAGACTCACCAGAAGCACTATGTCCAGCTCAGAATGAACT
GCTCAGTAAGCAGCCTTGTCAATGAGGAGGCAGCAGGCCAGCCCCAGAGG
CCTCAAAGTGGGAGAGTAGAGAAGCGCAGTTCCTGCCACAAAGGCACAGT
GGACACCTTGCTCCCCTGGCTGGCTGGAAGCAGATGGTGTCCACCTGCTT
CCATGGGAATTCTGCACCTTTAATAAAGTTTTATGGGACAGGAAGGTGAC
TGGCATTGACATTGTAACGAGGAATGGGTGGTGCCACCTTTGCTGTGTCT
TACCAGAAATACCTGTGGCAGGTAAATTTCTAGAGAGACCCTCCCATTTC
TCCCATATAGCAATTTTGAAATGTTTCCTGAGGGCTTTCCAAATTCATCT
GGGAACATAGGAGTTCCAGAAAGATGAAATCAAAGGTGATGGTATGCCAA
AGAAAGTAGCTTTTAGAATGACTTACATTAGCCATTCATCCATTCAGCAC

FIG. 3 (28 of 52)

ACCAGGCATTCAGTTTGAGGGGTGTGTGTGTGTGTGTGCGCGCGCGCGTG
CGTGCATGAGTGCATGCGCGCGCGTGTACATAGGGGAAGGGAAACAAAAC
AAAAGTACACAAGACATGATAGTTGTCCTCAAGGAGTTTTTGCAAATGTT
CACAATTTAAGAGAATATGCTGTGCTGTGGCTGGTGTATAAACCAACTGC
TAGGGAGAGGCCTTCCACACACACTTGGGGCAAATGCGACCTCTAGGACT
GCCAGTGGAATCTGGGCATGCTGTTTGTGGTCGATAAACCCTGGTCCCTT
GATCAGGGACCTATGTTTACTTTTCCTCTCCCTGGAAGTCTTCATTAGTG
GGCATCCAGAAGGTCTTGCACAGGGCAGAGGGAGGCACAAAGACAAGAGT
TTGAAACCAGCCTGGACAACAAAATGAGTTTCTATCTTTACAAAAAAAAT
TTTTAAAAAATTAGCCAGGTAGGATTGCATGTGCCTGTAGTCCCAGCTAT
TCAGGAAGCTGAGGCAGGAGGATTCCCTGAGACCAGGAATTTTGAGGCTG
CAGTGAGCTATTAAGTTGGCGCAAAAGTAATCGTGGTTTTTATCATTAAA
AGTAATGGCAAAACTTTTAATGACAAAAACCGTGATTACTTTTGCACCAA
TTTAATATGATTGCACGACTGCACTGTGCTCCAGCCTGGGCAACAGAGTG
GGACCCTGTCACAAAATAATAAATAAATAAAATGTAAACATGTAAAAAAA
ACCCCAAAAACAAAAAAAATGGGTGTTGAGACCCCTGAATTGAGGAATAA
TAGGAAGGAGTGTGATTCTGTGTGTGCATGCATGGGTGTGCACCCCTCAGT
GCCTGGGTGGCTTACCCTGGGCTAGTTCAGGTGGCAAATGGTTTTCCTCC
AGCTGGGCTACCACCATCTTCCCCCAGGGCCTGTCCATGTATTTGGTGGC
AAGATACCTATGGACTAGAGTCCCTCCTCAGAGGAAAGGCTCCTCCCATT
TCTCTGGCTTTCAGGTAGTAGTCCATGACTTCAACAGGTCCCCACTGCAA
TGTTATGGGTTAGTTTAGGTGGGGTCTCCTCTGAGAGCCTCCCATAGCCC
AAAAGGCCCTGTCCTAGCTGGCACTGCATCTCCCTCTTCCCAGCTCTCAG
CCTTTCTCTTTGCTCATCCCACTCCGCACAGGCTTTCTGCCTGATCCTTG
GATGTGTCAATCCTGCCCCTAAGGGATGCAAGGCAATTTGTCCTTTTATT
ATTAAGATCTCTCCTGAGGCCACGTGTGGTGGCTCACACCTGTAGTCCTA
GAACTTTGGTAGGCCAAGGTAGGAGAATTGCTTGAGCTCAGGAGTTCCAG
GCTGTAGTGAACCATGATTGCACCATTGCATTCCAGCCTGTGTGACACAG
CGAGACCCTGTCTTTTTTCTTTTTTTTTTTGAGACAGGGTCTCGCTCTGT
CATCCAGGCTAGAGTGCAGCGGTGTTTTTCTGCTCACTGCAGCCTCAACC
TGCACATTTTTTGTAGAGACGGTGTCTTGCTATGTTGCCCAGAGTGGCCT
CAAACTCCTGGGCTCAAGAGATCTTTCCACCTCAGCCTTCCAAAGTGCTG
GGACTACAGGCGTGAGCTACCGCGCCCAACAAAGACCCTGTCTTAAAAAG
AAAACAAAAATAAACAACTCCCTCAAGTCTTTTTTTTTTTTTGAGACGG
AGTCTCGCTCTGTCGCCCAGGCTGGAGGGCAGTGGCGCAATCTTGGCTCA
CTGCAAGCTCTGCCTCCCGGGTTCACGCCATTCTCTTGCCTCAGCCTCCC
GAGTAGCTGGGACTACAGGTGCCCGCCACCACGCCTGGCTAATATTTTGT
ATTTTTAGTAGAGATGGGGTTTCACTGCGTTAGCCAGGATGGTCTTGATC
TCCTCACCTTGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTAC
AGGCATGAGCCACCGCGCCCAGCCAGACCTCTTGAGTCTTAAACTCCTCT
GTAGTTCCAGCCACCCTTTAGCACATGACTCTGTTAATTTTGTTCTCACT
GTCTGAAATCATCTCCTGTCCACTCTTGACTGACAGGTCTCTGCACTAGC
CCACTGCTTAATCAGAGTAGGTCCCTGTCAACTTATTCATATTGTGTCCC
CATGCCAGTGTGGATGATTAAAATTGTTGAGTGGAGGCTGATCAGATGAG
CCATCTCCTTCCAAGTCCTCACTTGCTGGCTCCTGTCTTAGTTTTAGTCC
CCATTCTTCAAAGAACGTGAGCCCTGGAAAGTATTTTAGTCATTTAGTTC
AGTGCCTTTGGATGGGAGGATCACATCCCTGGGTCCCGTCCTGCAGACTG
TTTTGCTCTAGCTGACTAGGCAGGATTCCCTGCCTTCTCTCACTTCGGCA
TGGGACTTCCTTCTGAAATTGCTGCTCAGTCAAGAGAATGACCTTCCCCA
ACATAATCCTACTCCACAGGGACTTAAAGGTGTGTCAGAGATCTCTTGCT
CATCTTTCTGGCCAGGTGCCAACGTCAGTTTATAGCCAAGGGACAAGACT
AGTTAGCAGATCAGGCAGGTCTTAGACCCCAGCGTAAGTGCCAGACTTCT
AGCTGCAGTTGTTCCTGCCCACACTGGGCGTTCAGGTGGAGAGAGGGCAT
GGCACTACACTGAGCTCTCGGCGAAACCCAGGACTCTGAAATCTCGGTGT
CAGCCACAGGCCACTCTTTTCAGCAGGACTTCAGTCAGTCCTGTCACTAG
GCTGTCGAGCACATGGTAGGCTTTACCCC
>Contig35
AAGGAGTGTGCTTGCTGATAGCATGTGTGANGGGACGAGGAGTAAATAAT
TTCTGCCTTCAAGAAATTGCAAACTAGTAATGGAGATAAAATCAACAGAG
GAACAATTAGAGTATAAGGTAAAATCTAAGGGCCATAAGAGAGGAGAAGA

FIG. 3 (29 of 52)

AGTATGGGAGTTCAGAGGTAGGGGGTAAATGAGGGGAGTAGGTGGGTAGA
AAAGGTTAAAAGTAAATAATGATGGGAAGGAAGACAAAAAGACGACAGGG
GTGCCAAAGGACTCTTAACCTCATCTGAACGGAGTTGCCCTGTTTTGCTC
TCTGATGCTCATGTATCTATCCTTAGAGACAGCTTGGCGGGCAATGTAGA
GCGTAGGGGCTGACATAGGGGGTTGGAGTCCCACCTCCGTGACTTCTAGC
AAATTAGCAAACTTTGCTGCTGCTAAGCCTATAAGGCGGACAGAAATGCC
ATCTTTAAAGCTTGTTATGTAAAGTGCCTAGGACCTCGTAGGCATCAACA
GGAATAATGGATGAAACAAAACAACGGTGCGTATCTTGGAGAAAGTGGCA
TCTGAGCAGGAGTATTTTGAAAGGTAGGAAAGGGCTCCAAGCACATCTAA
GAGATTAGGGAACGCAGAAGCCTTAGCCCTGGGTGCAGATTTAACCAATC
AACTTCTAACCACCGCAGGCTGAGAGGTGTGGAGTGAGAGCCCCGCCAGA
GGCAGGAGACCCGGGCTTCGGCCAGACCCCGCCTCCTGGTACAGAGGACC
ACGCCCGGCTCTGCCTGGAGCCAAATGTGGATCAAAACAGCGCGCAGCTT
CCCACTGCTGGTGAAAACCCGAGCAAGGGGCCTCAGTTTCTTTATCCGGA
ACGTGGTGACAATGACATCTCTTTGCAAGGCTGCTGCAGGGCTTTCTGGA
AATACGCCCGTGAGGTATCTGGGCCTGCGCACAGCCTCCCCCGCCCAGGA
CCCAGACGTCTACCTGGGGGTCCCGTCTGCGCTCCCGGGATGGAAAACGC
CCAGGGGAAACTTAGGCAGGCGAGCGGACGGGCACCTCCCGCGGGACGAA
CTCACTCGGTGGCCTCCTACTTCCCCGGCCGTGTTCCAACGCCTGAGAAT
AACGGGAACAGCGGTCGTACTCACCGACAGCGGCAGCAGCGGTAGGCCCG
GGCCCCACCATGACTCTTCAGTGACAGTTTTTCTTCAAACGCCGCGCCTG
TAGCCAGGACCGGCGTGCCGCGCGTCCACGCGTCCTCATTGGCTCCTGCG
GGTTTGAAACTCGCTAGTCGTCAGCACGGGAGGGCGGGACAACAGGCAAT
AGGCTCTTTGCGGTTGGCTCTGGCCTTGAGAACCCGACCTTGGGGCCCTT
TGATTGGAAGAACGTGCAGCGCACCTCGGCATTGAGGGCGGCTTCCTCGG
GGCGCGGCGCCGCCCGCCTCTGAGTGCGCCTGTGAGTGCGCCTCCGAGTG
GGCGTGGGACCCTCCGTGGGGGCCTCAGCCGGGCTGGTGGTTGGGGGGCG
GTTACGCTGAATCCAGCTGGGGTTGGCGCGCCGGGAGTCCCTGGGCGGAG
AGACAGGGCGGTCCTCCCAGGATGCTGGGGCCGCTACCTGATTCTGTCCT
TTCAAAGTCTCAGACTCACAGGAGCTGTGAAAAAATAATATTATAAAGAG
GACATATGGGTCTTATGCATCTAAAGGCTCCTAGTTCTTAGTACTGCAGG
GTGGCTCGTTTAATTGTGGTAAAATATGCATAACATCACATATACCATTT
TAACCATTTTAAAGTGTTAAATTTTTCAAAAATGTGCAGTTTAGTGGTAT
TAAGTACCCTCACATTGTGGCACAGCCACCACTACTGTCCTTTCCAGAAC
TTTTTCATCTTCCCAAATGAAACCCTGTACCCGTCACTAACTCCGCACTC
CTCCCTCCCCCAGCCCCAGGCAATCACCATTCTAGTTTCTGTCTCTATGG
ATTTGACAACTGTAGGTGCCATATAAGTAGAATCATGCAGTATTTGTTCT
GTGACTGGCTTGTTTCACTTAGCATAAAGTATTCAAGGTTCATCCATGTG
TAGCATGTGTCAGAATTTCCTTTCCTTTTAAGGGGGAATAGCATTTCGTT
GTGTGGAGATGCCACATTTTGCTTCTTGGTCCATCCCTCTCCGGACACTT
GAGTTGCTTCCACTTTTTGGCTATTGTGAATAATAATATGAACATGAATG
CACAAATAACTCTTTGAGACTCTCCTTTTCATTCTTTTGGGTATATACCA
CGAAGTGGTATTGTTGGATCAAACGGCAATTCTATTTTTAATTTTTTGAG
AAACTGCCTTACTCCTCTCACGGTGATCTCTTGTTCAAGGTATATTTTCG
ATTTCACCTGATCAGCTGACTATAAGGCCATAAGGCTAACGGAGAAACGC
AGGCCTAGTTTCTCCTAGTTACTAGGAGATCGCAGGCCTCGTTGTCCTGA
ATCCCTAGACACACTTCATTCCCCTTGTTTTAATCCTAAATTTTTTTTCT
TTTGAAGTTTGTCCTGTTTCATCTATTCTCCAGTTTCTTAAAGAGGTCTG
GAAAATGCTTTTGGCTCCTTGTGTATGAAGGTTCCTCTTCCATGGATGCT
GGAGAAGTCGTGTGTGGAGGGGCAGTCATATCTGGGCACCTGTTGGCCAG
GTTCAGCTTACCAGTTGGGTACTCAGCAGGGCATGAAGCCACTGCAGCAG
CCCTTCTCTTTAGCCGTAAATAGGGAGTTTGGAAGAGAGCCAGGGTTTCT
GGATTTATGCATTTTGATATTTTCAATAGTGTATTAAATGTTTAAAATAG
GAAAACTGATCATTATTTTTGTTAATGACTGAGAAAGGGACTCCTTCACC
AACAGTTTCAGAAAAGTGAAGGCGGTTTTGTTTTGGTCTTTGTAGAATCT
AGGTGGTTGAATGCATGTCAGTTGTAGAAGTCACCTTGCCTGATATCCCA
CGCAGTGCTGGAGTATTCCACAGACCCCATGTAGGTACTGCACCTTTGCA
GGTATACTGCTGGTGTTGGTGAGCTGCCTTACCTGTCCTGTTATTGGAGA
CCCCTGCTTATTAGGAAACTTAAAATGAACTCAAATGAGCTTCCTTGCTT
ACTGGTCCTAGTCCTTTGGAGCAACATAGGCCAGTTCTGCCTCGTTTTTT

FIG. 3 (30 of 52)

TCCATCCTTTGGGTATTTGACGGTCTATTTTGTAGGACACAAAATGTGGG
AAAATAGCTAGGCAGGTTTAAAAATTCTCAACTCTACCAAGCATGGTGGC
TTATGTCTGTAATCAATCCCAGCACTTTGTGAAGCTGAGGCAAGAGGATT
GCTTGAGCCTAGGAGTTTGAGACCAGACTGGGCAACATAGCAAGACCTCG
TTTCTTAAAAAAAAAAAAAAAAATTACAAAAATTAACCAGGCATGGTGGCA
CACACCTGTAGTCCCTTCTACTCAGGAGGCTGAGGTGGGAGGATCACTTG
AGCCCAAAAGTTGAAGGATGCAGTGCACTGTGGTCATGCCACCGCACTCC
AGCATGGGAGGCAGAGCAAGACCCTGTCTCCAAATAAATACATAAATTAA
ATTCTTAACTCATTCATCAAAGTATCCACTGTAGCTTTCCATCATCCTGG
TGTTGTTTTTTTTAGAAGGATCTGGCTCCATTGCCCGGCTAGAGTGCAGT
GGCATGATCTCAGCTCACTGCAGCCCCCACCTCTCTGGCTTAAGCGATCA
CCCACTTCAGTCACCCATCTGGGTAATTTTTGTATTTTTTGTAGAGATGG
GGTTTTGCCATGTTGCCCCAGGTTGGTCTTGAACTCCTGGCTCAAGCGAT
CCATCTGCCTCCATCTCCTAAAGTGTTGGGATTACAGGTGTGAGCCACCA
CACCAGGACAATCCTGGTGGCTTTTAACGGTTTTCCATTGCTCTCAGGCT
AATGACCTATAAGCCCCTGCGGGCTTGGCCTTTACTCCCTCAGCATTAG
CCACCTCCCTTAGCCTTAGCCCACACTACTCTCCCCTTGCTCAGTGTTAT
CCAGACACTTTGTTTTTTCCTTTCCATACTCCTCTCTGTCTGGGAATCCA
ACCTTTCTTTCTCATTTCTCTAGTTGATTATTATTATTTTTACTCTAGCA
GCCTTATTGAGATATTTACATACCGTACGATTCTCCCACTTACAGTGTAC
AATTCAATTTTCTAACATTTTCATCACCCCCTAAAGAAACCCTATACTCA
TTAGCAGTCACTCCCCATTCTCCCCTCCTCTCAGCCCCTAGAAACCATGA
ATCTACTATCCATCTCTATAGATTTGCCTTCTGGACATTTCATATGTATG
AAATTATGCAATTTGTGGTCTCTGATGGGCTTCTTTTGTTACCAAAATAT
CATGGGTTTGATCTAGGTCCTGCTGCTCGCTGCACAGAAAGCCAGCCACT
GAGATGACAAGTATTGCCAAGGAAGAAGGCTTTAGTCAGGTGCTGCAGCT
GAGGAGATGGGGGCTCAATCTCAAATCCATCTCGCTGACCTAAAACCAGG
GGTTTGGATAGCAGGGAAGAAATGTAACAATGCGTAAGAAAACAGGAACC
AGGGAGGGGCAAGGAAGCAATCCTGATGAATGAGTGGTCCAAAGTCTCAT
TGCCTGGATGTGGTGATCTGGCGAGTTTCAGTTCTTTGATACTTTTTTTG
AGAGGCCTGAAGTCTTTTCCCCAGGAAGGAACTCAAACAAAACAAATACA
AGCTTCCAGCTTTAAGACCAGAAGCGTCAATTTCTATGTTTATCCGAAAG
AACAGTCTATGGGACTATTGGTTAAGTTTCACTTTCACTTAGTATGCTGT
TTTCAAGGTTTATCCACATAGCATGTGTCAGTACTTCATTCTTTTATGAC
TGGGTATTCTATTGTGCGGATATACAATATTTTATTTGCCATTCATCAGT
TGATGGACATCTAGGTTCTTTCCACTTTTTGGCTATTATGAATAATGCTG
TTATGAACTTTCATGTATAAGTTTTTGTGTAGACATATGTTTTCAACACT
CATGGGTATATACCTAATGAGAGGAATTACTGTGTCATACGATAATTCTA
TCTTTAACCATTTGAGGAACTGCCAGACTGTTTTCCAAAGCAGCTGCAGC
ATTTTACATTCCTACCAGCAGTGTATGAAAGTTCCAGTTTCTTTACATCC
TCAACAACACTTGTTATTGTCCATCTTTTAAATTACAACCATCCTAGTGG
TTGTGAAATGGTATCACATTGTGGTTTTTATTTGTATTTCCTTGATGACT
AATGATGTTAAGCATCTTTTTATGTGTTTACTGGCCATTTGTATATCTCT
ATTCAGAGTCTTTGCCAATTTTTAAATTGGGTCAGTTGTCTTCTTCCTTT
TTTTTTGAGATGGAGCCTCACTCTGTTTCCCAGCTGGAATACAGTGGTGT
GATCTCAGCTCACTGCAACTTCCACCTCCTGTGTTCAAGTGATTCTGGTG
CCTCAGCCTCCCAAGTAGCTGGGATTACACGCACCTGCCACCATTCCCAG
CTAATTTTTTTCTTTGTATTTTGAGTAGAGACGGGGTTTCACCATGTTGG
CCAGGCTAGTCTCTTTGTTGACTCTTAACCATCCTTCAGTCTCAGACAAA
ACATCCCTTTCTCAAGGATTGTGATTAGCTTGATTATTTGCTTATCTTTC
TCCCTGCTAGTCTGTAAACTGAGGGTAGGCCACTATATTCATTGTTCTTG
GCACCAAATAGAAACTAAATTAATGTCTTTTGAATGAATAGGGCTTTCTC
CTTTTAAAGATCCCTTCAATACAGTAACCACACTATATATAAGTAGCCAC
AAGCCCATTCAATAATACTACTAGTNCTTGCGCCAAACC

>Contig36

GGCTCAGCGTTACTATACTGGTCTCAAACTCCTGGGCTCAAGCGATCTGC
CCCCCTCGGCTTCCCAAAGTGTTGGGATTATAGGCGTGAGCCACGGTGCC
TGGCCTCAAATAACTATTTAAGTGAAACAAAACTAGTATGGCACTAATGA
AAAATGTATAAATCCATAATCGCAGAGGGATTTCAACTTACTTCTTTCGA
TTATGTAAAGGTCAAACAGACAAAAGACAATGACAAAACTTAATGCAATG

FIG. 3 (31 of 52)

AACACTTTTGATTTAATGAACATATATTGGATATGTACCCAAGAATTAGA
GAATACATACTAGTTTTGAGTTTATGCAGAACATTTACAAAAATTTAGTG
GAAGCCTAAATTATAAAAAGTTGCTGTCACGTAGAATAACACACAAACCC
CTGAGTCCGGAATTCAAAGCCCTCCACACTCTCCTCTACCTTTGCATCTT
TATCCTCCACCACACTGCAGTGCATACTCTGGGCTACTACTCACTGTTCT
TGATTCAAATTCCATGTTCTGTCAGCTCAAATCATTCTCTCTGCCTGGAA
TAACTACTTCATACATATTCTGCTATTGAATTCTTGTCTTAGCACCCCAT
CTACTCCAAGACGATGTCCAGTTGGGGTTACTCCCTGTCCCATTTTCTTT
GATTACACTTTTTTTTTCTACTTCCATTATATTATTGATCACATCTGTGC
CACAGTTTTTGACTTTGTGTCTGCTTTTACTCTTTTCTAGACCCTGATAG
CTCCTGAAGGGTTGGGTCATTTCTTTTTTATTTGCTCATTCCTCATGGCA
CAGTGAGTGCTTAATAAATGGCTATTGACTGAAATTAAACTGTATCTAAA
TGGACATATTCCACTTCTGGGCCATTCATTCTTTCTTTCTATTGGAACCA
GGAGATGGGGAACCATAACAAAGGTAAGGTTGTGCCATGTGAAAGAACAT
GGAACCTTCCCCTGAGGGCCAAAAAAGAGCAGGGAAAGGTGCAAAGACAA
AATCTTCCATTTTTAAACAATGTAAGAATGTGGTCCACCTCATGCTCAGG
TGGGACTTTATCATGACGTTATTTTTGGGGACTTATAGCTGCATCATTTA
CCCCATATACATTTACCTTTAGTGTAGGGAACTGAGGACAGGAATTTTGT
TGATGCAGACTCTTGCTAATGAGGCTAACACTTGGAGAATTTTTATCATG
CATTCAAGAAGCTTGTTTTACATTTCTTCATTAATACTTTAGTTGGTGGT
TTAGCTTTAGTTGTAGGCTTATCAGATATTTGGAGATATCTTCATAAACG
ATGGCTTTGGTTTTAGAAGAGTTATTCTGAAGCTACTATTTCTGGCAATA
ATCAAACAGCATGGCCATTTGTTTTGTAAGGCCTTTCCTAGAATATGACG
GTAAAATCTACGTGTGGAAAAATGCTTATTCTTCTGTCCTCTATAAATGT
GAATCTAGTTTGTCTTCAAAATGAAATCAAGTGATTAAAATGTAGTTTTC
TAAGAAGATAAATGGAGCAAAGCACTCTGTGTTTCACAGTGTTGGAAATC
ACTCATCCCTCATAAAACTGTCCCAACTGATCCTGACTCACATGAATGAA
TTAAAATAAGAGTTAATAACATCAATTTACATTTTTAAAGACACTTTCCC
ATGTTTTAGACTATTGGTTGGAAAAGCTGGTAGGTGTACAATTTGTGGAG
AGTTGGCTGTTTTTGTCTGTCGTTGTTTGACGTATTTCAAAGCCATATCT
AATTTTGTTGCAGAATGGTCTGAATTCTACAAAAATGTTGAGTTGTGTAG
TGTGGAGAAGTACGGAGCCATTTACTGAAAGGCTGGGGGGAAATGACGAG
ACCCTGAGATAAGGCAGTAGTGGTGCGAACAGAGTGGAAGGGAGGTAGTT
GAGATATGTTCAGAGTAGAATCAGAATGGACATAGTGAACAACTGGATGC
AGGTGGGGGCTGAGGAAGCAAAGTTGAGGATAATTCTGAGACTTCTAGGT
TGATCCACTGAAGTTACATTATTCAACACCACAAGGAAACTAGGGGAATG
AGAAGGCATACTGGTTTGCTTTGGAGTGGAAGGGCAGTGATGTAAGAGGA
GTTAATGAGTTAAAGTTTGGATATGCCTGAACTTCAATTTGATATGTGCA
TCTGATATACCCTTGGGGTGACCCTCCAGGCAATGGTTGAACATGTGTAT
TTCTTAGTAACTGATAGGCATCACAGACTCACATCAGTAAGGAAGCAACA
GCAAACTTGATTGGACGATATACCTGGAACTCAGTACCCTATGACTGGAG
CAAGTCTCTGTCAGTGAAATGAGGATAAGAAGAATCTTGACCTTGTGGAA
TATGTTGTTAGGAATATATGTGATGAACAACATAGGATACTTCCTACAGG
GCTCCACATGTAGTAAGGGCTTTATAAATGCTTGATAAATATTATTGTTG
TAATTTATTTCCAAAGTAAGATGCCACTGGAGGAATCTTTGGAACCCAAA
TTAATAACAAATAGGACTGGATGCAATGGCTCACACCTGTAATCCCAGCA
CTTTGGAAGGCCAAGGCAGGAGGATCTCTTGAGCCCAGAAATTCAAGACC
AGCCTGGGTGACACAGGGAGACCTTGTATCTATGAAGAATTAAAAAAAAT
TAACCAGATGTGGTGGTGCACGCCTATAGTCCCTGCTGCTTGAGAGGCTG
AGGTGGGAGGATTGCTTGAGCCCATGAGGTTGAGGCTGCAGTGAGCCATA
ATTGTGCCACCACACTCCAGACTGGGTGACAGAGTGAGACCCTATCTCAA
ATAAATAAATAAATAAATAAATAAATAAGTACAAACCAGCAAACACTAAT
CCTTTCTAGAGATTATTGAACTCTGGAGGGCAGATCTGAATGGAGCCAGC
AGAGGGACCTATGGAGATCAGCCTGGCCCTGGACAGCACCAGGCAATGGG
GTTGCTAGAGAGGTAATGGGGTTGAACAGGGTTTAAGCCATGAGGTCTCA
AGAATCCGTGAAGACTCAGACTAATTTTTTTTTTTTTTGCATGAGGATTAG
GTGTTCCTAGGAATTTCAATGAGAGCAGGGTTAATGAAGGAATGCAGGGT
AGGAGAGCTGAGGGAAGGCATCTGAGAGAGCCTGGCTTATGAATGGCTGC
GTCAGTATGGCTCACCTGCTTTCCTTGTATCTACTTAGCAGATGATCCCA
CCCCAGGCCTCCAGGGCCAAGGTCATTTCCACATAGTCATGGCCCTTGA

FIG. 3 (32 of 52)

GGGCCTGGAGCAGTGTAAGGAAGACAGAGTCTTAAGAAATTGCATTAAC
GTCATGGTGCTTGGCAAGTGTCGTCATCCTATGCCAAGCCTGATCTGAAG
GGGTGCATGCTCATAGGTAGCTGCTGCCCAAGATTACAGCAGCTTCTTCA
ATCCCAGATCCATGCTCTCCTATATTCATTTTTCCAGGGGTTCCTGTCCT
TCGACAGTGATGAGATGCAGAATGACTTATTGAGTTATTCTCCTGATAGT
TGCCAACTTTTCCAAATGACAATGGGGCATGGAGCTTGAGAGTGGAAATG
AGGCCCTAGGGATAGCGTGCTTAGGAAAACACTCCCAGCCTGATGTAATT
CTGGGGGTACAATGGCATTTTCATCATCAAGACTGATGTAAAGGGTGACT
AGCAGTGAGTTGGGGGTGACTCGCACTGGGGCTAGGTTTCTGATTCTGCC
TAATCCAGACAGAGCAGAAGCACTAGTGGGCTGGTAGAGGGCCTCCAGGG
CCTCACTTAATGTCCTGGAAAAACAGCTCCAGATTGTTGGTTCACGTTCT
GAGGACAAGCTTGGGTACTACAGGATAGAGAGAGTGGTGGGAGATGCCGT
GGCCTGCCCTGCTGATGCCTGCCCTGCCATTCCTGCGTGTGATGTCTCTG
GGGCATCTTGCCTTCCCTGCCCAGACCTGTAGTTCAGCTGAGGGCATGTG
GAGGCCAAATGGCTTCTTAGAGTGTTACTTTCCTTGAACAGCTCTGCTGG
GAGAACTGGAGGAGCTAGCTAGTCACGGTAACTGCAGCAGTCAAAGGATC
GTCCCGGTGGAGGTGGGGTGGAAAGGTAGAGAAAGAGAACATATAGCGTT
TTCCTTGGAGATGTGTGGGCATGTCATAGAGGAAATACCCAATTCCTGAG
CCTTGAGCCCTCCAGGAAACCTTGGAATATTAGGTTAGTCATCCCCAAGG
AAGTCTAAGAATTCTGGTCTCACCCATCTCCTTTAATTCCCACAATGATC
CTACATGATATTAAGGAACACGGGCCAGTAACCCTCCAAGCAATGGATGT
GGTGGTGAAGTTTGACCTCATGATGGAGCGGAGGTTGGTTTGAAACCTAA
GAATTTAATTTATTGTTTCAAACTGTTCTCCACTCAGCGTTATTAAAGCA
TACATAATTGACACATAAAAATTGTATATGTCTACGGTGTACAATGTGAT
GTTTCGATCTATGTATACATTGTGAAATGATTACAACAAGCTAAATAACA
TACCCATTCATCGTGTTTCAAAGGAATTAAACTCAAGCACAAAAGAGAGG
TGCTGTTGAAGAGTAGGGCTGCTCTATCTAAGTAGTATGTCTGGGGTTGT
CCTGGATCAGGGTCCTTTTGTGCTAGTAATAAACCAGCCCTTCTGGGGCT
GCTCCACTTTCCCCACATTTTCTTCTGGAGCCTCCCTAAGAATTAGGACA
TGGCCACTTTCTCTGCATAGGCTTCCTACTTCAACAAGGACAGGGCTTGT
GCTGCCCCATGCCACTTGAGTGTCCCTACAGCACAGAGCTGAGTGCACAC
TGGCTGAGTGAGGAAATCCCCCAGATTAATCTTGGTTCTAAGCATCATGG
CTGTATTTCACACGTATATGAATTACAAATTACAGCATAGTCGAATAAGG
ATTTTTGTGCTACAACTGGAATCCCAGATTATGCAAATTGGATAGTATAA
TATTGAAATTCCTAGGACTTTTTATTAGTTTTAAAAAATTATACAAGCTT
AGAGTAAGAAATTAAACAGTGCAAAAGAATTCACTGTGAAAAGTAAAATG
CTCTGTCTCTGCTGAGAGACAGATATTGCAGCCCAGATACTACTGGGGTC
AATAGTTTCCTTTAAGCATGCCATTTTGATGGTTTATGGGACTTACAGCT
CAAGAAGCTTGACACTAGGGTTGATCTCAGAAAATCATTGTTGCAGGTAT
TAGATATGACCGTCTCATAAAGATACACACACAGACACAGCGATTGGAGA
TATTCACTGGGGCTTATGGGCTGCTTGTCCTTTCTGCTCTGTGCCTAAGT
TGGGCTCAGAGTAGCCTGGCATCGGCTGTGGGGAGAATGCTGGCATGGGG
TTAGCAGGAGCCCACTTAACATGTCCTAAGCCACCTGGAAGAGTCCTTCA
AGGAGACCAGACTCCAGAGGCCCTAAGGAAGGAAGGACTTTTGCCCGTTT
TTAGGTATTCTAGTCCCAGAGTTTAGGGAGGAATGGTTTGGCTTTGGGTC
GTGTGCCCCTTTACCGAGTGGGATGGGATGTGCCCATGAGCTGTTGAGCT
GGCTCTTGGAGAAGACAGCAAAAGCGGGAATAAGAGGTCAGGAAGCTGTG
TGGTTGTAGGAAATCCCAGCAGAGGGCCTGGGGGTCAAAAGTGGTCATGG
TAGTGACGGTGGAGGCTGAGGTGGTAGAAAATCAGAGGACAAACCCCATG
GGCTGCTGGTGATCTGACCGAGCTCCTATGCTCTCCTGGTTCATTTTAGG
CTCTGTAGCAGCAGATGATTGGCTGGTGTGAGAGCAGTGCACCTGCCATA
TCAGGCAATCCAAGACAAGTCCAAGCTACGCTGGGAGGAAACCTGAAGGC
AGCAGCAGGTAGACTGGCTGAAGACAGACAGGCAGGCAACTTGTCAATCA
GATTTGTGTTTTTAAGGACTTTTAACTGGGGAGCCCTCCGGGACAGATCA
GATGAGAGTGAAATGTGCTCCGCCTTAGCC
>Contig37
GGCCGTTCGCAATTCTGTAAAAGGGAGAGTGGTTTTATTTATTTTTAAAC
ATAGTCAAGCTGCTAAAGTATATGATATGTATAGATAGAGTATAATTAAA
TACTTTCAACTACAGACAAAATCAGGAGAATGGAATTAAAAAACAATTTA
CAAATGGGTAATGGCAGCATTGGGTTGCGCCCACCCACGAGAAGGCAGAC

FIG. 3 (33 of 52)

ACCAAGATTCTAAGATCACACGTGGCCAGCACTTCAGACTTCAAATAGAA
TTCGTGATTATGCATTATTTTTCTCGGAAAGTTTTCACTTCACTATATGC
TACTTGACACTTGCTTTCCTAAGACATCCCTCTATTTTTGAGATGACTAA
CTCAGCAATTCATTTCTCTCACGCATAAGCTGTCACTCAACCCAAACCCA
CCAAGCCTGCATTCTACCCTCAATAAGGTCTTGGTGTGTAAACTGACCCA
CTTCACCTAGTTCCTTAGCCCTCTCTTGACCAGACATGACTCTTTCATAA
GCTAGACCTATAAAGTCAGGGCTCTTAAGTAGCTGATCTCTGATAGTGCC
AAGTGTCCCCCACTGTTCACATTTTCCACTCCAGCTTCTAACAGGTGATA
GACTGCTTTTTGGGGGTAGGGGCACCAAAACATATAGACCTCATGTTTGG
ATGTAGACACTCCAGTTTCTTTAAATTACAACTACATATTAATAATGACT
TCCAAGTGTACATTTCAGTCCAGATCTCTCCCTGGATCCCCAAACTTTGT
AAAACCCACCGCCTAGTTGATATCTTTTGATGTCTGACAGGCATTTCAAA
TTTAATACTGTCACAAACAAAGTTATTGATTTTCATCTCTGCATCTGTTA
CAAATTTTTCTTACTTTGGTAAATAGCACCCCAGGCTGTGTCACTGCCAA
GAACTTTCCACAGCTCTTGGAATAAAATTCAAAATATTTTCCAAGGCAGA
AAGGCACAGTGTAATCTGGCTCCTGCCTACCTCTCCAACCTCGTATCACA
CTAGTCTCCCTGTCACTCACCCCCTCCAGGAGCTCAGGTATCCTTAAAGT
TTCTTTTCTTTTTTTTTTTTTTTTTTTTTTGAAACAGTTTTGCTCTGTT
GCCCAGGCTGGAGTGAAGTGGCATGATCTCAGGTCACTGCAACCTCCGCC
TCCTGGGTTCAAGTGATTCTTGTGCCTCAGCCTCCCAAGTAGCTGCAATT
ACAGGCGCGTGCCACCACACCCGGCTAATTTTTGTATTTTTAGTAGAGAT
GGGGTTTCACAATGTTGGCTAAACCGGTCTCAAACTCCTGACCTCAAGTG
ATCTGACCACTTCAGCCTCCCAAGGTGCTGGGATTACAGGCGTGAACCAT
TGTACCCTGCCTCCTTGAAGTTTCTTGATCCAGACTCATTCCTGCCTTAA
GGTCTTGCATCTTCAGTCCTCCCCTCAAATGACACCTCCATGAAGACGCA
ATTACCTGTAATTACCGTGTCTTATTTAGTCAATGTGTTGGTTTTCTGTC
TCCTCCACTACAGTGTAAGCTCTATGAAGGCAGAAACCTTGGCAGTCCAG
TTCCCAGCACAGTGCCTAGCACACATAGGTATTTAATAACACACAGTAAA
ATTCACCTTTTAGTGTGCAATTCTGAGTTTGACAAATGCATCAAGTCAT
TTAAGTCTGACTATTATCAAGCTATAAGATGGTTGCAACACTATCACTAA
TTCCCTCATGCTCCTTGGTAGTCAGTCTCACCCCTAACGCCCCCCTCCTG
GCAATCACTGATCCGTTTTTTGTCTTTATAGTTTTGGTTTTTCCAGAATG
CCAATAACTAAGTTTTGAATGAATGAATGCTATTAACTCTCATTTCTGAC
TCCAGAGCAACATCCATGCAATATTTATTATTTCAGCCCCAAATACTGCC
CCCTCACCTTCACTCCAACCACCTACTTGATGATACAAGGTGAGACATTT
GGCATGTGCTTCCTCCATGTTCCTAGCATTTTCCCTATCTCCTTAGCCTT
CCTTCTAATCATAAACGAAGAGTGAACTTTCCCTTTCTAAAGGCAACTTA
CTCCTAGGACCTCGATGCCATAATTTTGTTTCTCTAGTACTTTCTATATA
TACACCAAACAATTAGCTCCAGAAAGGTAAAGACTCACTGTGTGCTCATC
ACTGTGTCTCCTAGCGCCTGGCACACTGCAGGTGCTGAAGAAACACCTAC
AGAATGAGTGAATGAATCTCTCCCTCTCTAGACTCCTTCTCTCTTTTGTAAT
CAAACATGTTCAACCTGCAACACAGTCTTATGACCAATCCTCTGTTGTCT
GACCTAGGCTGAGCTCCAGGGCTGGGACCCTGACTTCCTTATTCACCACC
TCAAGGTCTCTGCACTCACTTCTCTTTCTGCTCAGGATTGTTTTTCTTCT
TGTCACCAGTCTTTTCTCAGACTTAGGTCTCAGCTCAGACATTGCTGTTG
AAAGTACTTCTACTGATCCTTTTATCTAAAGCAGCCATTCCAGCCCTACT
CTCTTGATCATAGCACCCTGAATTAAGTTGTTTACTTACTGTCTCTTCAG
GAGGGCAAGGAGCTTGGTGGTGGTGTTCAGGGCTGTACCAAGCTGTACCT
TGCTTCACCCTGCTACACTTTTTAGCAACCATCTAATTTTACATGCTCCC
TTCACTCGTCAGAAATTTCCTTATTTTCTACTTCAAGCAGGTATACATAT
GTGCTTCTCCTGGGAGGCTCACCCACTTCATGAGACTACATTTGGTCCTG
GGTAGAAAGTGTACAAAATCCACTGGCTCAGTTTTAATCAATGTATGTTA
ATATTAACCAACCTGAGATCTTGATTTCCACGCCTGGCTAATTTTGTATT
TTTAGTAAAAACAGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAACTCC
CGACCTCAGGTGATCCGCTCACCTCGGCCTCCCAAAGTGCTGGGACTACA
GGCATGAGCCAGCGTGCCCGGCCTAAGATCTTGATTTCTACCATCTGAAC
TCTGTATTTGAACTGACTGCTCCTGCTTGAGCTTACTGGCCAAAACTTGG
CCCACTCAGACTCACGGAAGTTTCTGGTTCTTCCCTGGTAACTTTTCTGA
ACTTAACCACTGGTTTGCTTGACAAGAGATTACCATCTTCTCACTTCCTA
GCTATGTGAACTCACTTATCTGCTCTATTGCTGTTCAGTCTAGCACGGCA

FIG. 3 (34 of 52)

CTTATTGAACGAGTGTCTACATCTGCACCCCCTACTTCTTACTCATCCAT
TCTGTTTCAATTTCTTAAAAAGAAAAAAAAAAAGCTATTGTAAACATACG
ATTACAGAAAATGATTTATAACATGTGTATGTACCACCTAGCCCTGTCAA
GTCTTAATATTTGTTATATTTGCTTCAAATCTTTTTTCAGACTGTAGTTA
AAAATTACTTAGGAGCCATTATTTATGGCCTATTTCCTGACCTAGTCTTC
TTGATGGTCAATTTGCCTAATCATCTTAAGTTGCAAAAGCTTAGAATTAA
AGÇAAAGTACCTTCGATCCTCTGCTGTTGCCTTCTTTTTAATATTTGGGT
TTGTTTGGGTCCCATTTACGGTTGTGACATCAGCTTGAGTTTTGGGAGCT
GTCTTGTTCAGAAAATGGTTCTGGGGAACAGCCTTTTTCAACTTGGAGTC
CAAAGTCTGTGCTTTTTGCTGAAAGCCATTATTGTTATGTTTATTACCAC
TGGTTCCATTTGGTCTTATGCTAGGGGTGCTTGGAATGGCTGAATTAAAT
CTGCCAACTGTCAAATTAGGCCTCTGGCTTACGGCTTTTGACTTTTGCAG
TACACATGATGTCTGAGGTATACAAACTTGGCTGGACTTCTGATCTTGCT
TGATGTTTGGATGTCTGTTGTTATATTCACCCTGAAGCAAACTGGGGTAT
GTTCTGGGTTTGGTGTGCTTCACTCTCTGTTCAGTAACAGGGTATGACCG
TATCTTAGTTTCATTTGGTCTTTCATATTGACTCCTATTAACCTTTATAT
CTTTGATGTTCTTGACTACTGGTTTCTTTGATGACTGAACTTTACTAAGG
GTCCGAATAAAGTGAGAGGGAACCGTCCTTGAGGGTTTTACTCCTGGTCT
TGCAAGATCTGCTCCTCTAGAGAGTTGCTGTGATTTTACTGGGAAAGTCC
TGCTTTGTGTTTCTCCAACAAATTGTTTATTAACCCTATCTTTCAGAACA
GCACTATTAACTGAACTTTTGCCCAAGGCTTGTTTAGGAACTAAACTGTT
CTTGGTTTGATTATAAGAGTCAGTCTTTGGCTTACTTCTGGTATATAATT
TAGGATCTGGCTTCCTCTCAGGTTCTGTTAAGATATCTAGCAAGTTCTCT
TTGTTTGTTTCTTTTAGAAAGTTATCCAAAGATTCGTTTTCAACATGGAT
ATTATTCATAAAGTCTATACATTTACCATTTCCTTGATCTGTTAACTGCT
GCTTTGTAGTTTTCAATTGCTCTATATTAAGTGACCCCACAGGTTTTCTT
GACAGTCCTCCTGTGGTGGACTATCTAGCTTCACACTGTTGAAAACTCTT
GCTGAAAAGCTTAGACTATGGGTTAGAAGAAACACATTTTGAAGTCCGCC
TTTTGCCCAGAAGTTTTGGTGGCTCTAACTTCAGCTTCTGGGACCCTGCA
GTATTAGGTGGTCTGGGCTGGAGTTTAATGCTGATGGACCTTTTAGGTTT
GACAGGCAAAACAACATGGTTGGTAACATCATTTTTGGGTCTAATAGTCT
GAAAAAACAAAGAAAATACATATTAAAAAATCCTTAACATATCTTATTGT
TTTTAAAATAATAACTGTGTTTAACACATGCTAAAAAAAAAATCATTTTT
AGAATTTCATCTAAGAAAGTTGAATCCTCAGAAAGTAAAGAAAGACTCAC
TAATAGGTAGTTTTTGTGTTTTTTTTTTTTTTTTTTTTTGAGACAGGATC
TTGCTCTGTCACCCAGTCTGGTGTGCAGTGATGCAATCTTGGCTCATTGC
AACCTCTGCCTCCTGGGTTGAAGCAATTCTCCCACCCCAACCTCGCAAGT
GGCTGGACTACAGGCGCATGTCACTACACCTGGCTACTTTTTTGTATTTT
TAGTAAAGTTGGGGTTTCACCATATTGGCCAGGTTGGTCTTGAAATCCTG
ACCTCCAGTGATCCACGCACCTTGGCCTCCCAAAGTGCTGGGATAACAGG
TATGAGCCACCACACCTGTCCTAACAGGTAGTTTTTACAACTTGAGTTCC
TATCAGAAGTATATTAGAATCTTTTAGCTTGACAGAATTAAGCAGAGATG
CAGTGAATATACAAAACTTGCTCTTTCAAAAATGAATTTGCCTCAAACAG
TAGTTGTTGAATGCCTATTATATCCTAAGTGCCCCTCCAAAGAACCCTGAA
AAAATACATACATAATGAACTTATGTTAGGGTACCTCCCAACAAATCTCT
CCTAGTACTTTGTATAGCCACACTATATGTTTTTTAAACCACTGCCTTTG
TAAACATCACAGTATCACTCAAGAACCTCTGTCTCATCCCTGGAGATCAG
TGACAAGGAGATAGGTGGCAGATGATGTGAGGCCTGAGATATGCTGCCAC
AGCTCTCAATAAACATGTAACATCTTAATAGTCATATTTGTAAAATCAGC
CAGGACAGGGTTTTAAGGTTAGAGTCTATGTTAATAATAAACAAATGTTT
AGTCATGTGATTTAAGTTTGGATAAGAAAGGTAGGACTCGATTACAGAGA
ATTTTGAAAACTAGGGAAGGGAGTTTAGAATTCATATGGTAAGTAATTGG
GCAAGCCACTATGAATTCCTGAGCATCTCTCATGAAAGCAATTACTCAGA
AAGGAGAATTTCACAGAGATTTATGGAATATGTTTCCAGGGTAAGATATG
GGAATGCTAGAGTTACCACTCTATTTTTGATTTGACAAATATTGTGAAGA
ATCACTACATAAACTTGGCGAGTATGTAAAGGATTTCTAACCAGAACCAT
TTGGCATTGAGGGCAAAGAAATGTCTACTCTGGATGATAGCGGTGTGTGT
GGTGTTACTAGGAGTGAAACAGCGGAGTTGGGAGTGGGAGGCAGAGAGAT
GGATGGTATACCCACAATGGCTATATCTGGATTAATCTTTGAGCACCAAC
ATTTATATACACCTCGGATCTCTCCATCATTGCTTACTGAAGAGGTGGAG

FIG. 3 (35 of 52)

GGACGTTGGCATGAAAGCTTCCAAATGTGTTTTTTAGTTGCTTTCTTAT
ATATTAAAAACGAATTGATATAATCCACAAACCATAAAATTCACCATTTT
AGTAAGTGCACACTTCTGTGGATTTTAGTATAGCCACACTATTATACAGC
AATCACCACTGTCTAATTCCAGAACATATTCATCACCCCTAGAAAGAGAC
TTGGGTTTACTTGTTGGCAGTCCCTCCCCA
>Contig38
GGTCTACATGTGCTCGCAAGATTGGATATTGAAATATCAGCAAGAAATTA
AATGACATAGTAGTCATTATGCCTAAATTATTGTTATTTTTTGATTGAAA
AAAGTTGAATATTTCAAATATCAAGGTAGTAGTGAGATATAATAAAGAGA
GAGTCAGTTCTAAGTATAGAATTGCTGATTCAGTTAAGCTCTGTTCTCCA
ACATTTGGGCCACATTGAAGAGACCATGTAGCTGCTTTCAGCCTCGGTTT
CCTCCTTTGCAAAATGGGGATTACACTACCTGCCTCACAGAGATGTAAAC
TTATGACATGTTATCATGATTGCCAGGGCCCACCTGTTTTCTTTTAAACA
TTGAAATCACTGTGCCTGAAACAGGGATTTCCCTGCCCTTTGTGCAAGCT
CCAGAAACAGGAGTCAGCCTGAGTCCCGCAGCTAAGAACGTGGATTCTGG
TCATTTTCTCATAGCGAACACACTTCACAGGTCCTTCAAGGGAGTACATT
TTCCTATAACTCACCTTAATCTCAGTTGAAGCCTCGTTTCTTATTTTGCA
CTGTGGCCAAAAACTAAATCTCATTTCTTTCACGTAAACTTCAGCAATTC
AATAATAGTACAGTCATTTTATGTTTCAACTGAACCAAGTCAGGGTTCCA
CTCCTGCCTCCCCTTTCTGCTCTGAGGACATCCATGAAGTGGAGGGGGTC
TATGTAGCCTGGAGCTATTGGTGAGGGGCGATGGGTCCGTGGTGGTCTTG
GGGAACTGCGGGGCTGTGTCTGGCTGGTCTGGTGTCTGGTGATTGGCCTT
GTTCCACGCGGTTCACGCTGCAGGACAGTTCGTGTCCTTCTTGTCCTAAT
GATCAGCTTTTAGGCTCACGGGCCTGTCTCTGCTGAGATATGGAATAGGA
CAGCCTCTGGATCTTCTTTAAACTCTCCTGGGGCCACAGGGGACTCTGTT
TGTGTCTGTGCCCACATAGGATGATTCTGCCCAGACCTTTGCTGCCATTT
CTTGCTGTTCTGCTGTTTTTAGTCTCTGGAGGGCTTGCAGTTTCCTTGGG
GTCCCTGTGGAAGCAAAGCAAAGTCCTCTCCACGCTCAGATGTCTAAACG
TATCTGGGTTTTATCGTCCACCCATCCCAGAGCTCAGTCTAGAGGAGGGG
GCAGCCTTCGGGTTCTCTCCTTCCTCCCAGAGCCTCTTCCTTTGCACCAG
GGCAGCCTCTTCCTATCTGTTGGAAAGGGCTGTCTGGTTCTTGAATATAG
AGTTGCAGGTTTGAGGGGTGTAGGCTGAGGTAAGGCAAACTATCACATGG
AATAAAAATTACCCTGTGTCAAGGAACAACCAGAGCTGGACAGTTTTTAA
ATGTGAAAACCAATTTTATTCAGGACTATGGCGAGAGGTGAAGTAAGACC
TCAGTATAGAACTGGGCTCAATTCCGAATGCAGCATGGGCAAATGGGAAT
GTATAGCCTAGGAGCAGGGTGGGAACCTGTGGATGAAGAATTACTAAAAG
GGCATATCAGGGGTGAGGGGGCGTCCTGGCTACACCCACTAACTACTGTT
GCTGAAGAAAGGCCTGGTGACATCACTGGGGAATGGTGGGGATGAAGAA
TCCAATCAGATGGATATTGAGGATAAGGGGATCTTGATAAACTGGCTTAG
GAGGGTTTTTGCTAAAACTGGTTTTCATAGGTAAGTCCACAGACAGGTCT
TGGAGAAAGTTCAGGGACCTACGGTTTGTTCGGGCAGATGCTTTGTCATC
TGTCACACTGGCACTGTCACCTGGCTTTCCTTTAGTCCCTCCCCCCCTTT
TTTTTTTCTGGAGTAGTTTTGGGAGACCAGAGGAGCAGGGAGTTAGGGAG
AGTAGTCAGAAAAGGCCAGAGAAAATAAGGAGGTGTCTGTAGGGAAAATC
CTTAAATCCTCTAATTAAATTAATTTAATTTATTTATCTGGACAAGGTC
TCACTCTGTTGCCCAGGCTGAAGTGCAGTGGTGTGATCTCGGCTCACTGC
AGCCTCGACCTCAGGGCTCAAGCAGTTTTGCCACCTCAGCCTCCTGAGTA
GCTGGGGCTCACAGGTGTGCACTACCATGCCCGGGTAATTTTTGGGTTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGAGATGAGGTTTCGCCATG
TTGCCCAGGCTTGGTCTCGAACTCCTAAGTGATCCATCCACGTCGACCTC
CCAAAGTGCTGAGATTACAGGCATGAGCCACTGTGCCCGGCCTAAATTCT
CCAATTTTTAAATGCTTCCCTGTTCCCTGTTCCAGATTTGGGATATTGAC
TGCTGTTAAATCAGCGATTTCTCCCTGTGGAGAGGTAGCCAATAGGAAGC
AACAAGAGTGAGGAGTCCTTATATCGAAATAGAGGGTAAGAGAAGAGACA
GATGTTATCTTGGCAGTGATTTAAGAACAGCGAGTCTGTAAGCAAAGCAA
AGCAAGGCTCCCAGGTGCTGAGAAACAATGGCTTTCTGGGGAAGCGTCTG
TGTTCAGAACCTTAAGTTGGAAACATCTCTGAAGATGTTTGCCATGAAGG
TTTTCTTCTGAAGTTGAGTCTTTCATCACTAGGTAGGCGTGTTTTGGAGT
CTCTATCAAACAGATCCTGTGTTTATTAGGAAGCTGTGGTTCATAAAGCC
CCATGCTAATTTTGCAGGTAGCAGGGTGGCCCTGGCCTGACCCGGGGACA

GAGTGGCTGTCCTCCCTGCAGGCAGGAAACTCTCTCCTGCCACCTAGTGG
CTGCATACCCACATTTCAAGGGAGCTTCTGGGTGGTGAGTTTACCAGACT
ATGGTCTGAGGTAGAGTTAAGCAAAACAAAACTAAACTGCATAAAGAAAC
AGAAAGAAAATCAGGTGTTATAAAAACAATTTGGCATTTGTTTGTGTTTC
AGCTCCGTGTCGATTTATTGCTTCCACAAATAGTGCCGATATGCACCAGG
CACTGTTGTAAAACTGAAAATATGTTTTTGGATGTGCCCAGTCTGTGAGT
ATTAAACGATGGTTGATTTGAAATTTGCTATGATTCATATTTCTGGGGGT
AAGATGCAGGATTTCTTTGGGGGGCCTACGATGTGGCATTCTAGAATTCT
CAAAGAATCAACCCTGGTGGGACCAGGAAGAGCTGAGCTGAGGCCTCTCT
GCTCATGTGTACTTACTGGAGATCATGGAGACAGGTGAGCCTGAGTGCAC
GTCTCACCAAAGCCACAGCAGAGGGGGAGGAGGCGGAAAGAGAGCTCTCT
CCATTTCTGAGAAGTTAATGGTAACAATGGCATACATACCTACTTTACAG
TTGAAATTGGAAACCACAGCATTAAGTGTTTCCAATGAAATTTGGCAATT
TGGGAGTTTTCTGAGCTGCATTGGATGTGGTTTTGCATGCTGTTAGGATG
AGCAAGAGATGATGGAGAACATCTTCCTTTTGAGCTTCCTCTTGGACGTG
GGTCACTCCCACTCATGGAATTAGAAAGCTTAGACCTAGACTTGAATCTC
ACCTTCTCAAGGTGCTCCCGGGCAAATCACTTAAGATCCATCTTCTTCTC
CTCCTGCTCCTTCTCCTCCTTCTGAGTTTTTTTTTTTCTTTCCAAAATTC
AAATGACACGGTACTGGTAGAAGAAAAGGTCCAAGTCTGCTTTTACAGCT
CCCCTCATCCCCAAATGTACTCCGACCCCAAGATGACCATGTTATCATTT
GATTGACATCCTTCTAGTTTCAACTCATTTCTTTGCATGTATATGCACGT
ACATATACACTATTTTATTTTGCCAGGGGTCACCGTTTAGCTGCATTAAT
TTCTTATAAAATAATCTATATTTACTTATGGTTTACGTAAAACAACATAC
ACATGTAAGTGTATAGCTTGATAAGTCTTCACTGTAAACCAAAAATAAAA
TTCGAAGCCCCCCCAACCGTCTGAATGGACCCCTCTTCTTGGCCAAGAGC
ATTCCAAAGTTAACCTGAAAAAACTAGTTCAGGTCATGATGGAAGGGAAG
GTTGGACATGCCCCAGTATACCCTTCTCCCTTTTGGAATTCAGGAAAAGC
TGACCAGCATTAACATCAACACAGACCTTATGTCTGATAGGAAACTTTGA
CAATCTATTCCCTCTGAAGCTTGCTACCCGGAGGCTTCATCTACAAGATA
AAACCTTGGTCTCCACAACCGCTTATCATAACCCAGACATTCCTTTCTGT
TGAGAATAATTTACCTTGTAACCTGGAAGCTCCCTGCTTCAAGTTCCCTC
ACCTTTCCAGATTGAACCAATGTAAACCTTACATGCATTGATTGATGTAT
TATGTCTCCCTAAGATGAATAAAAGCAAGCTGTATGTTGACTGCCTTCAG
CACAGGTTGTCAGGACCTCCTGAGGCTGGGTCACGGATGCATCCTTAACC
TTGGCAAAATAAACTGTCTAGATTGACTGAGACCTATCTCAGATACTGTT
GGGTTCAAATATATAACTTATGAAACTAATACACAAATCAAGTCATAGAA
TATTTCCATCACTCCTCATCTACCCCCAAATTTCCTTATGCGTCTTTGCA
GTCAACCTCCCACCCCATCCCCAGGCAACTGCAGATCTACTTTTTGTCTC
TGCACCTTCAACTGACCCTTTCTGTGATTTCATATGAATGGAATCATGCG
CTGAGCAGTCTTTTGTGTCTGGCTTCTTTTGCTCAGCATAATGTTTTTGA
GGTTTGTCCATGTTTTTGTGTTTGTCAATGGTTAATTTCTCTCCATTGCA
GAGTAGTTTTCTATTGTACATGTGTACCACAATTTGTATATCCATTCCAT
TGCTGATGGACATTTGATTTGTTTCCAGATTTTGGCAATTATGAATAGAG
CTACCATGAACACCCAGGTACAAGTCTTTGTGTGGACTTATGTTTTCATT
TCTCTTGGAATGGAACTGTCATATCAATAAGTATATGTTTAACTTTGTAA
GAAACTGACAACAAATTATCTGCGATGGTTATGCCATTTTGTTTTTCTAC
CAGCAATACACGAGCATTTCAGTTGCTCCACAACTTTGCCAAAACTTGTT
TTCTTTAATTTGGACATTTAAGTGGTGTACAGAGGCATCTCATTGTGGTT
CTAGTTTTCTTTGCCCTGATGACCAATGGTGTTGAACATCTTTTCATGTG
CTTTTTGACCATTTACATATCCTCTTTTGTGAAGTGTCTGTTCAAATATT
TTTGCCCATTTAAAACATTTGGGGGTTTGTCTTATTATTGTGTTGGGAGA
GTTCCATATTTATTTATTTATTGAGATGGAGTCTCACTCTGTTGCCCAGG
CTAGAGTGCAGTGGCGTGATCTTGGCTCACTGCAACCTCCACTTCCTGGG
TTCAAGCAATTCTCCTGCCTTAGCCTCCTGAGTAGCTGGGATTACAGGCA
TGTGCCACCACACTGGCTAAGTTTTTGTATTTTTAGTAGAGATGGGGTTT
CATCATGTTGGCCAGACTGGTCGCAAATTCCTGACCTCAAGCAATCCACC
TGCCTCGGCCCTACAAAGTGCTGGGATTACAAGCATGAGCCACTGTGCCT
GGCCCATATTTATTTTTTATTCTTTATTTTGTATACAAGTTCTTGGTCAG
ATACAATAATACCTGGTCAGATGAGATAATGAGTTGGAAAATGCTTTGCA
AATGGGGGAGAATAATTTAAATGTTATTTATTTATTAAGAGCAGAGGCCC

FIG. 3 (37 of 52)

TTCCTGTTGCGGTCACAGAAGCCGTTTGCTTCTTCTGCCTTTTATAAACC
AGCAGAGTCGAGCTACACAGGCTGTCTGTGTTGGCTGCTATTAGTTAATC
AGAGAGTTTTTTTTTTCTTGCCTTGTCATTCTAATTTGTGACACATAATT
AGCCACAATATGTGTTTTCAGTTGTGACACTGGCCTGGGAAACCAAGGGA
TGTTTAGAGTGGATTTCCTTGATTTGCAATAATTGTGTGTTTTTCTGCA
TCTTCTGTTAAACACAAATTCATGGAAGCAAAACATGGAAGCAAAGTACC
CTGGACATCCCCCCTTCTTTATGAAATTGATTTCTCTTAAATGTAATGTT
TGCTTGTTCCCTTACTTTAAAAGCAATTTAAGAGTTTATTGAGAAAGTGA
GCCCTGGAAACATAGATGCATAGAGAGAAAATTCTACCACCCTCAGGTCC
CTATTGTCTTCTCTCATAAAGTGTAGTTTCAGGGCCTTTTAGAAGTTTCT
TTTCTGCTCTGATTTGCATGTTTGTGAGTGTTGCTATTTTAAGTATTTGG
ATTTGGTCTGCAAATCCTATGAGAGATGGCAACAGAGTAGGGATCTCAAA
GCCTGCAGGTTGTATTAAGTCCAGCAGGGCCTTGTATTTACAACAGAGGG
TCCTTGAAGACATTCCATATATTATGCTAGGGGAGTGGCCAAGCAAACTT
TAATGTGTCCCTATGGTGGGATATTTGGGGTTAATACCTGCCCTTCTCTT
AATTTCTTTTTCTTTTCTTTTTTTCTTTTTCTTTCTTTTTTTTTTTGAAA
TGTAGTCTTGCTTTGTCACCCANGCTGGATTGGAGTGCAGTGGTATGATC
TCAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAATTCTCCTGCCTC
AGCCTCCCAAGTAGCTGGGACTATAGGCACACACCACCATGCCTGGCTAG
TTTTTTTTTTTTTTTTTGAAACNGAATCTCGCTCTGTCGCCCAGGCGGGA
CTGCGGACTGCAGTGGCGCAATCTCGG
>Contig39
CGCTCGCATCCCTCATATCCATGAGTGTTCTGTGGGCCCTGCCTCTGAAA
TAAATCCTGCCTTTGTCTCCCAGTTCACTCCAGCCACCCATCCTGGGGCT
GCACCCTCCTCCTTCCAAGCCCTCTCCCTTTCCTTCCTGGTGCTGCCTGT
CATGTCAAGCATATGCATCAGTGCGACCAGGACATTTGAAATGCAACCAG
TACAATTGGGCGCGGTTATGCCTACCAGTTTTTCTTCCTTAAACATTTTA
TATTTATGTTTGAAAGCATGCCACCTTTCTTCACTTGCCAACTTGACAGA
TTTATTAGTTGACAACATCCGCTGATAGCATCAGTAATAAGTTAATTGTT
TTTGCACATGTAGCTTTAATTATTCTCATTATCATTTATAGGAGTTATTC
TTTGTAAAGGGTAACTGAGTTTTCCAAAACAAACAGAAATTTGGGGTGGG
CCCATGGAGCGTGACTCATGAAATCAGATTCTTAGAAGGACCTCGGCAAG
TCTCTGGGTTGCTGTTAATGAGCCTGGCTGGCTGCCAGGGGTGTGTCTGC
CCTTTATGAGGCCACCACTGTTCAAATGCTTGCCTGCAGCATTACTTGCC
TAGGTAGTGCTTGTTTCTACTGAACTGTCAGGGATCCAATTCTTTGTGGT
CTAAGTAACAATACTCAGATTCACAAGGAATTGATTAATAAGCCAGAATG
CCAATGTATTACATTTTTGATGAAGACCATATTTACAGTGATTGTATCTG
CTCAAGCTCAAATTAGGATTAGAGTTCTGACAAATACATATGTGAGAAGT
ATGAGGTTAAATACTTGAAATTTGGACTTTTCTAGAAAATCTGAATGTGA
TTGCCATTCACATACCTTTCTGGGGATGATGATTCTTGTACTTTTATTTT
AAAAGACATAGAAAACTAACTTAAGAATCAGATTGCTTGGCTGGGCACAG
TGGCTCATGCCTGTAATGCCAGCACTTTGGGAGGCCAAGGTGAGTGGATT
GCTTGAGCTCAGGAGTTTGAGATCAGCCTGGGCAACATGGTGAAATCCCA
TCTCTACCAAAAATACAAAAAAAAAAAAAAAAACAACCAAAAAGAATAAA
TTAGCTAGGTGTGATGGTGCGTGCTTGTAGTTCCAGCTACTTGGGAGGAT
GAGGTGGAAGAATTGCTTGAGCCCAGGAGGTGGAGGTTTCAGTGAGCTGG
GGTTGCAACAGTGTACTCCAGCCTGGGCGATAGAGTGAGACTCCGTCTCA
AAAAAAAAAAAAATCAGATTGCTTTATTGCTGGTTTTCTTTCTAAAACTGA
GATTGGGTCCCATCATCCCCTGGCCCCCATTGGTTAATGGTTCCTCCTTT
GTCTATTGAATAAAATACAGATGTCTGCTTTTGGCAACATGGTTGAATGT
AGACACTGCAGGGTCTTCCTGACTCAAAATGAGTAAGGCTTAGATAAAAC
ACATTTTGAAATGCATTTCTGGATGAACAGCAAGGAAAGGAGATCTCTTA
AAATCCTCTTTCTGTTCCCCTCTCCCTACCCCCTCCAAGTGGGCTTAAGT
AGGAAGGGTGGTGAGCGGCAGGTAAACACACGTCAAAGGCAGTCTTCCTC
TCTGAGGGAAAACACTTGTATAAGCATTGCAATCAATGGGCCTCTTTAAT
TATGTGCCAGTGGCAAGAGCGGGTGCTGAACCCAGGGGCCTGCCTCAATC
CGGGGCCTTTGAGGCAGAATAAAGTGGTCTCAGGTTGTTGGCATTTCCTT
GCCCTTCCACCCGAAGCAGACACAAATCCTCTCTGGAGGCAAGTTCCCCA
ATTCAGCCAGTACAACTCCCACAGACTAAGATCAATCATGTACAAGCTCA
CAGACAAAGGTCACCAAACACACAGAGCAATAAACAAATTCATGAGTGAC

FIG. 3 (38 of 52)

GTGAATGAGAATAAACA_.AACAATAACCACCAGCTGGGATGCTCTAAG.
CTTCAGCTGTTAGAATTCCTGAATATAGAATAAAACTGCCACAATGGCAA
ACATGCATCTAGTACTTACTGTGTGCTGGGTTCTAAGAATTTTGCACATT
GTGCCAGATACCGACTCAGCTTCACACTCACCCTCCTACTGTGCCCTCTT
AATTTGCACTAGATTAAAAGGTAGAAAGGAAGAGGCAGCTATTCTGTTCT
TGGCTGTGCCTCTGGCAGCACATGCAAAATGGGCAGTAACAGTGGCAGTC
ACAGGTAAGTAGCCTTCTCACAGTGTGGAGTTAAAGGCATGGGACTGAGA
CGAGCAAGGTTCCTAAAGGGACAGTGGCCAGTAGATGACCAGGGGCTACT
GGAGTGGCTGCATGGCTCTGTGGAAGCTCAGAGGAGCCTTGGGTCCTGCA
GGTGCAGTAGCAGCTTTCTGTAGTTCCTGATCTCTGGGTCCCACAATCTT
CCCCGTTTTTGCTCCTCCACTTCTAATTTTGTAACTGACTTCCCTGTGTG
TACTTCTCTCTCTGATTGAAATAGCCAGACTGGTTTCTGTTTCCTGATAA
GACATTGTCTGGTACGAACACAGTAACTCATTTAATCCGATATCTCTATG
AAGGAGGTACAATAATTATTCCTATTTTACAGATGAGGAAACACAGCAGA
GAAATAAAGTCAATTGTCTAAGGTTGCACATTTAGTCAAGGGAAGGGTTG
ATATAACATATAATTATTTAGAAAACATCTAAGGAAATAAAAGGCATAAT
TTAAAAATAAAACTAGGCAGGTTTAAAAAAATGAAGTAATCTATAAGTAA
AAAAGTATAATTGTTGAAATACATATCTTAGTGGATGGGTTAAATAGCTG
AAGAAATGATTAATGAACTGGAAGGTAGTTCTGAGGAAATCAGAATTCAG
CATAGATAGAAAAAAATGGGAATTTACAAAAGTACACAGGAATTATAAAAG
AGGTTAAATTATAGGGAGGGTAGAATGAGAATTAACATTGGTCTAACTGG
AATTTTGGAAGAAGAGAATAGAGAGAATGAACAAGGCAATATTTAAAGAG
GTGGCTGAGAATTTTTCAGAACCAACACAAACTATGACTTTACCAGTAGA
GAAAACAATGTACACTGAGGAGGATAAATAAATATACTATGAACAAATTG
TAATAATAATACTCAACAAAGACAAAGAGAAGATGTTAAAATCAGCAAAA
AAAGAAAGTCAGACTTAGAAAGAAATGACAATGGCAGACTACTCAACAAC
AACAATGGAATCCAAATTCGGTCAAACAGTATTTTCTTCATGCTAGCATA
TAGC
>Contig40
GGGAGTCCGCTATGCTCCTAAAGATTTGCACCTCTGATCTGGTTTGTAGT
TAGTCTCTTTTATTGCTTTATCCTACTCAACTAATTTTTTTAGTGCCTGT
TTTTTTTTTTTTTTAATGTGTGTTGATGACTACAATTCTAAACTCATTCTA
CTGATTCATGGGTGCTTTAAAATCTGAGCAGTCTTTCGCATTTACTGCCT
GTGATGGCCCATCCCACCAGCTAAAGTGTGTGGCCACTGCTTACAGCACC
ATGTGATAACGAGTAAGGGAGAGATGCCGCCCAGACTCTTCTAGGAGCAG
CCAGTAGGACCTTCCAGGGGTTGCAAGCAAACCACAGCAATATGTGGAGT
GTGGCAGAGGATGGCCCCAAGAGGATGTGGCAGCGGCTAGTGCAGCTCAG
CTTAGTCTGAGAGGAAATGCTGGAGAGGAGAGCCCAGTCTGTACAGGCAT
GACAGCCACAAGGACTTCAACAGCTAACATGGCTGAGTGGACTTTATGTG
CTATCTCATTCAGAAAACAGGAGCAATCAGAAAGGAGTCACCTCCTATTT
GTACCCCAGGAATTGCTAACCTACTTGCATCTGAATGATGTCCATCACTT
CCCTTCATCACCTCCTCTGGGGGCTCTGCAAGGATTTGACTCCTGCATTA
GTGATCTGTCTCACCTACGTTGTGATTCACATGAACTTACTAATGTGCTA
TGTGACAACTACCATCTTAAACACAAAAACCCTCTTTTGATTCTGTGGCT
CCCTCCAGCTACCCCTGCATTTCTCTGTCCCCCTGCCCCGTCTCTGCACT
CACTTTTATTTTACAGCAAAACTACTCAAGGGAGTCTCAGTGCTCCTTGG
CTCCATGTCTCCACCTTTCATTCTCTCCTCAGTTCACTCCTGTCAGGCTT
CCGTCCTCAAGCTCTTCTTCACTTTTGTTCTAGGGCCGCTGACATCCTCT
TTCTTGCCAAATTCAGTGGCCAGGTCCTCACTTACTCAACTGCTCAGCAT
TGTTGGGCCTGGTGGACCACATTCTCCTTCACCCACCTTTTGCTGCTCTC
TCTTCTCTCCAGATGTTTCTCTCTTCTCACTGGCTACTCCTCTTTTGTCT
CCTTTGTTAGCTCCATTTCTTCCTTCCAACCTCACTGTGCTGGTGTGCCC
AGTGCTCAGTTTTTAGCTATTCTCTCTTTTCCAGTGGCATTCATTAGATG
GTATCATGTGACCCATGGCATTATATGCCTTCTACATGACAGTTACTCCT
GAATATGAATCTCAGGAAAGATTTGGATTTATTTTTAATTAATTTTTTTA
AATTTTATTTTAATAAATGAGGTCTCTCTCTGTCATCCAGGCTGGAGTGT
AGTATTGAGTGATGTGATTATAGCTCACTGCAGCCTTGAACCATGGGCTC
AAGTGATCCTCCTGCCTCAGCTTCCTGAGTAGCTGGGACTACAGGCATGT
GCCACCATGCCTGGATGACTTTTTGTGTGTGTGTGTGTGTGTGGAGACAG
GGTCTTGCTCTATTGCCCAGGCTGATCACAAACTCCTGGCCTCAAGTGAT

CCTCTCACCTCAGCCT..CAAAGTGCTGGGATTACAGGTGTGAGACCA...
CTGGGCTAAGATTCAGATTTTGTATTCAATTGACTGTTTGACATCTTCAC
TTGGACACCTAAGAGGTATCTCAAATATTAATTAACTTGGCCAAAATACA
GAACTTTTGACCCCTGCCCCCACAATACTTGCCCCTTCCCCAGACTTCTC
CATTTCTGTTAAATATCCCCAGTTACTCAACCCTCAAACCTATGAATGCC
CTTTGATTTCTTTCTTTCCCTCATCTCCTACGTTGACGCCATCAGCTAGT
TTTGTTGCCTTTATGCCCAGAATATAATCCTCACCACCTTCTCTCCTATT
GCCCGAGTATAAGATGTCAGTTTTTCCTGCACAGTCCATTGCCCTGACCT
CCTGAGTGGTTTGCTTCCACTTTTGACATTTGTATTCCTCTTTCCCCCAG
GGTCAATTTTTCACAGCAAGAGTGGCATTTTTTTTTTTTTTTTTTTTTTG
AGACGGAGTCTCGCTCTGTCGCCCAGGCCGGACTGCGGACTGCAGTGGCG
CAATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCT
GCCTCAGCCTCCCGAGTAGCTGGGAATACAGGCGCCCGCCACCGCGCCCG
GCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCTTGTTAGCCAG
GATGGTCTCGATCTCCTGACCTCATGATCCACCCGCCTCGGCCTCCCAAA
GTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCAAGAGTGGCATTT
TTAAAACCATATATTAGATCATTGCTTTTGTGTTTGGGAACCTCCAAGGG
CTTTGCATCATATATCAAGTTGACACCTCTCCTACCCAAGCCTGGCTCTT
TCCTGCTCCTCTGTCCTCTCAGCCCCTCCACCCATTGTTCATGCTGCTTC
AGCCACACTGGCCTTCTTGCCATGCCACATTTGTGCTAAGCCCACATCCA
ATCTCGGGGCCTTTGCACTCGCATTTCCTCTGCTTGGCATGCTGTACCCC
AGATCTTTCATGATTGGCAGCTTCTGTACATTCAGCCACCTGCTCAAGCC
ACCCTTTCAGAGGGCCTTCCCTGGCCACCTCACCTGAAATAGCACCTCCG
ATTGCACCCATCCGGTTATTCTCCATCCTGTTCTCTTGCTTGGTGATTTT
CCATCACTGATGAGGAAATGAACCATGGAATGCTAGGGCTGATGACCAGA
ACTTTCCCCCACCCCCACATTATTACAGAGGAGGAAATGAGGTCGGAGGT
AAGATGGGCCCAGGATTTCTACTCCCGCCTGGACTGCAGGCACAGCACTG
ACCTCAGCTGTGCTCACTCTTGGCATTCACCCAACCCTTCTATCTCCAAC
TGCCCCATTTACCAGAAAGTGAAATGTTCTCAGAGACGGTGAGCCACCTG
ACTTGGACAGCAGCCCAGGGCCCCTGGCACCCTGCTTTCTTCCTCCCTGC
CATCCTTTCCTCTCCAAGACCTACCTTTCCCTGTGATTCTTGCCCACATG
CTGCATTTCATGGTTTTATGACCTGATTTCTGAGAGGGATTTGAATTTTC
ATGATTATTTATGTAAGCAAATCATTATGCTTATACAAATGAGAAAAGGA
GTGCTTCTGGACTTCCCAGGGACAAAATCTTGTCACTTGGCTTGCTTTCA
TATTGCTAATTAAGGACCCAGGATGTGGGTGAGATGTGCTAAAAGCTGAG
AGGAGGCTCTGGACTCTGACTATGGGCCCACACCCCTGGGCAGGCATCAC
ACTAGTCCTTTAGGTCATCCTCAACCCAGCTTCCAGTTGAATCAGATGTT
TGTGAATAACTCAGCAAGGCTGTATGGGAAATGAAGAATGAGGTGGGGAA
GAGGCCTGTGCAGAAGACACACTGACTTACCCCTCTACCTCTAACTAGGG
TGTTGTAGCAGCCACCCACCCACCAAGTCTGTCTTCCAGACCACGTATGC
TTTCCTCCACCTTTGCATCTTTTATCTTCTGCCAGCCCAGATGCTTGCTG
ACTCCAGCCCAAGCCTATAGGATAAGCTACAGCCTGTCCCTACAGACTAC
GCATTGCAGAATCTAAGACATCAAGTCAAGTTCGGAAGCACTTGCCTTCT
CCTCTCCAGGTACACAGGCTCTCCTGGAAAGCTGGTAGCAGCTGTGGAGG
TGTGGTGTGTTACCTGCTGCAGGTGCAGAGAAGTTGACTTCACAGCCCTT
CAGAAAGACTGCCTTCTTCCAGTTGTATTTGTGTACTTGCTTGGGTGTGG
GGAGGATTCTCAGCTTTCTCCACTCAAATTATCAGACCCTTTCCATTTAG
TGGTAGACCATTTCCCTCGTCCAGGCCAAGGGCACATAGTACAGAGAAAT
AGGGAGTTGTTACCCAGGGAGAGAACTTGGCTCTAAACCTGTAATAGAAA
GGTCAGTTCTGGTCTGGAGGGTCAATTTTGATCTTTGGCTCAGATCCAGG
AATTGGAACCAAGGCTTTTGAACATTTTAATGCAGGGGATTAAAAAAATG
ATACGAGTCATTCACGAATATATTTGCTTAACATCTAAAGAGATCCCTCA
AAACACTAGAAAAAATAAGAACAAAAATCTAATAAAACAAAATTTGTTAA
ACACATTTACCAAATTTTTTTTTTTGGTAAAAATTCAAATGTCATAAATA
AAGCTAAAGTTCCTCTTGATGACTCGCTCCTCTGCCCTATTCCACTCCAA
GTAACCACTATTATCAGTCTTGCCAATACCCTTCCAGACCTCTCTACCTC
TATATACCATTAGAAGCACATGGTTTTGCATTGAGGATGTGCAGTGTTTT
GTTTTACGTAAATGTTATCACTCTGTTCTTGTTCCATAATTTGCCTTTTT
CTCTCAATGATTTGCTTGGCTATCTTTCTATTTCAGTAGCATCTCCTTTC
TTTTTAACTTACCATTGTTTATTTAACCTTGCCTCTATCAACAGATATGT

FIG. 3 (40 of 52)

AGGTTGTTTCTAGTTGA.TTCATTAAGTATTTATAAACAACGCATCAGTA
GATGTCCATAAATTTCTTTACGGAAGATGGCAAGTAGTGGAATTGCTGAG
CCAAAGAACATGTTTAAAAAACCCAAAAAAACTAGACGCTACCAATTTTC
TCTCCAAAATGGCCATACCCACTTACCCATACAGAGATGATTTGGAATCT
GGCTTCCTCACAAGGTGAGATGCCTTCACAGTTTCATTCTTCCTGGCATG
TCTTCCTTTTGTATCTGAGAGAGCTGGCAGAATTGTGTCACTAAATCAA
GGATAGAGGGTCAAATGACAGCTCAAGCTCACAGGCACCTCTGCTTTCTT
CCCAGACCACCTGCTTTCCTGCCACCAGCTCTGTTCCATCTTATAGAATG
GTTGCCACTTGGGTGTCTGCTCCGACAGCCATGTCATCCTTTGCACTGCA
GTTATGAAGCAGACAGAGCTAGGAGAGGGGCTTTGCCAGCCTCTGCCCTA
GCTTGGAGAATTTCAAAGAAGGAGGGTATTGAGAGTGAGCTGCCGAAGAC
TGGCAGCTCCCTCAACTCAACAGTTGTCCTTCCACAAGAAGTCAGATACA
TTTTTTTGGGATAAAATATTTAAAAATTATTATTTTATTTCTGAATAATA
TATTTACATGATTCAAAAATCAAACTGTAGGCCAGGCATGGCTGCTTATG
CCTGTAATCCTAGCAATTTAGGAGGCCGAGGCGGGAGGATCACTTCAGCC
CAGGAGTTCAAGACCAGCCTGGGTAACATAGTGAGACCCTGTATCTACAA
AAATTTAAAAACAAAAATTAGTTGGGCATGGTGGCTGATATGGTTTGGCT
CTGTGACCCAACTCAAACCTCATGTTGAATTTTAATCCTCAATGTTGAGG
GAGGGTCCTGGTGGGAGGTGATTGGATCATGGGGGTGGGTTCTCCCTTGC
TGTTCTCATGATAGTGAGTGAGTTCTCACAAGACCTGGTTATTTGAAAGT
GTGTAGCACCTCCCCCTTCACTCTCTCACTCTCCTGCTCCGCCATAGTAA
GATGTGTGTGTTTCCCCTTTGCCTTCCGCCATGATTGTAAGTTTCCTGAA
GCCTCCCAGCTATGCTTCCTGTACAGCCTGTAGAACTGTGAATCAGTTAG
ACCTCTTTCTTCATAAATTACCCAGTCTCAGGTCATTCTTTATAGCAGT
GTGAGAGTGGATGAATATAGTGCCATATGTTTGTATTCCCAGCTACCCAG
GAGGCTGAGGTAAGAGGATTGCTTGAGCCTGGGAGTTTAAGGCTGCAGTG
AGCCATGACTGTACCACTGCTCTCCAGCCTGGGTGACAGCGAGACCTTGT
CTCCAAAAAAAAAAAAACCCAAACTGTGTAAAATGTGTTCATAAAAGTGTC
TTGCTCCCACACCTGTCCCTATATATCTTATTCCTCAGCCTCCGACAACT
ACTTTATTCATTTCTTATGTATCTTCCAGAATCAAAAAAAAAAAAATCAAA
TACAAGCACAGTGGAATGTATTGCCCTTCTTCCCCTCCCTTTTGTTACAT
CAGAGTTAGCATATCATAAATACGGTCTGCATTTTCTTCTTTTTCAGCTA
TCAGCATGTTTTGGAGAGGATTTCATATTCGTGCAGACAGCATGTATTAG
TCAGTCCTTGCATTGCTATAAGGAAATACCTGAGACTGCATAATTTATAA
AGAAAAGAGGTTTAATTGGCTCACAGCTTCGCAGGCTGTTCCACAGGAAG
CATGGCAGCATCTGCTTCTGGGGAGGCCTTAGGAAGCTTTTACTCATGCA
GAAGACAAAGCGGGAGTGGATGTCTTATATGGCAGGAGCAGGACTGAGAG
AGAGAGAGAGAGAGAGAAAGGATGCCACATACTTTTAAACAACCAGATCT
TGTGGGAACTCTGTCACGAGAACAGCACCAAAGGGATAGTGCTAAACCAT
TCATAAGAACTCCACCCCCATGATCCAATCACCCCACACCAGGCCCCACC
TCCAACATCGGGGATTACAATTTGACATGAGATTTGGGCTGGGACACAGA
ACCAAACAATACCAGAGTGCTTTCTCATTCTTTTCTATAGCTGCCTAGTA
TTCTATGTCCTTTACTTCATTTAGGCAGTCTCTTGTTGATAGACACTTGG
GTTACTTCCAATTTTTCCTATTACAAATGATGTGCAATGAATAATTTTGA
TCATTTTCCATTTCACATGGGTTATGTCCATCTGTGGGATAAATCTCCAG
GAGTGAAATTGCTGGATCAAAGGGGAAGTGCACTTGTGATTTTCATAGTT
AGCAAATTTTGTTCTATAAGGGTCATATCAATTTATAGTCCCACGCGTAA
TATTTAACAGTGGGGATTTCCCGACAGTTTGACCAACAAGGTCTGTTGTT
AAACTTTTGATTTTTGTCAATCTGATGGGAAAATACTAGTATCTCAAAGT
GCTTTTAATTTGACTTTCTTATTACAATGTTAAGCATCATTTTACTCTGC
CCAAGATCAAATAGTATTTTCTTTTCTGTGAACAGACTGTTAAGATCCCT
TGCCTCTTGTTTTGCTGGATTTTTGTTCTTTTTTTTCAAATGTTTTGAGG
CAGTTCTTTACATGTGAAACAAGTTATCTCTTTATCTGGGGTGTGAGTTA
CAACTACTTTTCCTCTGGCTTGTTTTGCGCTTTGACTTTGCTTCTGGTGA
TTCCCGCAATTCTGAAAGTGTACTTTTGCATCATTCATTCTTATACACC
CATGCTCTTGTTCACGCTGGTTCCTCTACCTGAGGGCTTTTTCTTTTCTG
CTTCTATCTGGGAACATTTTTTTGAGAGAGAGTCTCACTCTCTCGCCCAG
GCTGGAGTAGTGCAATGGCGCGATCTTAGCTCACTGCAACCTCCACCTCC
TGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACA
GGAGCCCACCACCAAGCCCAGCTAATTTGTTGATTTATTTATTTATTTT

FIG. 3 (41 of 52)

TGTAGAGATGGGAGTC.ACTATGTTGCCCAGGCTGGTCTTGAACTCC.G
GGCTCAAGCGATCCACCCACCTCGGCCACCCAAAGTGCTGGGATTACAGG
CGTAAGCCACCATGCCCAGCCCATGTGTGGAAATCTTCTGTTTATCCCTT
TAGGCTTGATTCTTATGTCGTTCTCCTCCCTCCTTCCTGGATACTCCTCT
TGTTCTTTATCTTACTCTACTTGTCATGTTACCTTGTTTCTGCTTATAAC
TAGCTGCCTCTCCTATCTGAGGAGGGACTTGTGACTGTTCTCATCTCTGT
ACTCCCAGCTCCTAGTACATAGCGCTTGCTCAACAGATGTTTGGTGCATT
GATAGATAAATCACTGGTAGCTGTTACTACCAGTCCTGACTCCCTGCAGT
GCTTCAGCTGATCCTGTTCCAGATGTGCACTGAATATCCTTCTGTTGAAC
AACAGAAATAAAGGGGATGGGTGAGGAGGATAGTCTTCGGTGGCCAAGGA
TATTTTTAGGTACTTTGCAGCACTCAGCAATGAGGAGTGGGCTTTAGTCC
CCCAAGAACTCTCACAGCCCTGGGTGTCTTTACTGTTCAGTGTCAAATCC
AAGACAAGTCAATGATCAGGAAAGACCATTTTTTTTTGTTCAGTGAAGTT
TATTTCAGAATCATTGAACAGTATGATATTTGGTAATTTCATAAATATTC
CCACTTAAAATGATCGGAGCAGATATATTTTCAGTCGTAATTAAAGGACA
TGATTTAAAGAGAGCACACCAGTCCAAATTGAAATGATTCCATAGCTATT
AAAAAACTAGGGTTTTTTACAGACAATGATACTTTTTGCCCCCTTTGAAT
AGATTAGACCAATGAATAAAACAAACAAACAAATAAATAAATAAATAGGG
AAGCGGTTGCTCATCAGAATGTGGGAGCGAATGACAGAGGGTTTCTTAGA
ACCAAATGTGGCCGTGGTTTCTGTCAGGCGTGCTTTAAGTGAGTAGGAGA
GGTGAGAGAGGCCTGGCTCAACAAAAGGGCTGGGGATTGTCCCTGAAGAA
CCAGAGCTGANTTNCATCAGGAGTAACANAGGTAGATAG
>Contig41
CCGCGTTGAGGTTCCACGCAGTTCAAATTATGTCCAATTATCAACATTAA
TGCACATTTTCAATAGAACCTGTTCCGGCTTTTCTTAGGAGGGGGGCGGG
GAGACGTTGTTCTCTGGGAATAAGTGTACGCAGGAGGCTGAGAAGGCTTC
ATTCCATAGCATTCACTTACCTCCAGCTGTAGAGTGGGCTTATCATCTTT
CAACACGCAGGACAGGTACAGATTTTTTTCTTTGAGGCCCAAGGCCACAG
GTATTTTGTCATTACTTTCTTCTCCTTGTACAAAGGACATGGAGAACACC
ACTGAAGAAAGAAGGGGGTCTTGTGGTTAGGGACACAGCAGTGCAGGGTC
ACCCCAACCCCTAGGCCCCATGAGTAGGATACATGTAATTTGGTAGCCTC
TGTGGGAACCCACAGTGAGGTTCCTTGGCCTAAGACACAGGATAACTTGA
CTTCTCACAGACAATAGCAGGGTCATTTTGTTGATTTAGGGTTTCCCCTC
AAAGGCCTGAGGGTTTCTCAGAGCCTCATAGCAGTAGGAACGGAGAATGA
AAGAGGGTCTACATTTTAAATGCTGAAGGAAGGAAGGAAGGAAGCCATTG
TGTCACTGGCTGGCAATGTGCCCATCCACAGGAGCGGAACAACTTGATCA
ATGTGGAAGGAAAGGAAAGAGGTGAGGCTGTACTTCTGCCAGAAATCAGG
CACCAGAACTGTTTCAGGAACAGAGAGTAGCCCATGGGAAGAAACTGGGA
GAGGAGAGGCTGAGCTGGGAAAGTGGCTCCAAAGAGAGACACTCATTTTG
ATCTTCCTCAGTCACAGCAGTGTCAATTGGAGGCCCTGGGATCACTCTTA
CTACCCGATTCCAAAGAAACAGGATTTTCTTGGCCTGGCTGAGAGCAAAT
AGCTTCCCCCTGAGTGAGGCTGTCCTTCAAAGTCAGCAGCCTTAGTTGCC
CACACTCCTGTGCAGAGGCTTTGGCTACTGTGGCACGATGCCAGGCAGAT
CACCACAGCTAATGATGGGTTCACCGCACTTGAAACTTTTGCCCGTTACA
GCGGAGAGATATAAGTTCCTGCTGGGCGGTAAAATTTCCCTACAAGGAAC
CACCTGGCATTGGGTGGACGGATGTTGGGGCAAGGGGGGAAGACTGGGG
AGGGGGATGGACACATTATCGCTCCAGCACTCTTGTTTCAGCCTCAACAA
CAGGAAGAGAGAACCCACAGGCAGTTAGGCCATGTCCATCAAATGACCCC
ATATTGTGGAAGAATTGACATTGCACTATGCCCAAGAGACTTGGGTGGAC
ATGGTCCTGGGAGTGCTTGAGCCGTCTAATTTCTCAGGGTCACACTCCTG
TTAACAAATGCACTGGCCAGTGCAATCAAATGTGCCATTTCTAGGACCAA
AGTTTGTATATTCCTTTTTAATATTTTTTTTCACTTGTGTTGATCATTTG
CCTTAAATTAACTTTCTACTTTGTTTAAAACATGGAGAATTAGCAAGCTG
CCAGGAGGCCAGGCAGGGAAACCAGGATGTTTCCATTTACCTTGTTGCTC
CATATCCTGTCCCTGGAGGTGGAGAGCTTTCAGTTCATATGGACCAGACA
TCACCAAGCTTTTTTGCTGTGAGTCCCGGAGCGTGCAGTTCAGTGATCGT
ACAGGTGCATCGTGCACATAAGCTTCGTTATCCCATGTGTCGAAGAAGAT
AGGTTCTGAAATGTGGAGCACATGTTTGTTTAGGTATAAAATCAGAAGGGC
AGGCCTCGTGAGGCGAGGTGGCAAAATTTGATTTCTTGGAGGACACCTGA
GCATATACGGTCAAAGTCTGATGACAACACCAGTAGGGATGAAGCTGGGA

FIG. 3 (42 of 52)

GTGGGGTGGCTAAGAAC CTGGACCTGACACTATTAGACATGGGTTCC
CTTCAGGTCTATTACTGCTCACTGTGGCCGAGCAACAGAGCTACTTAGGT
AAAATGGTGATGGTCATAACACTAGCCCACAGGGAGGTTACGAACCTCTG
GTGACAATGTAAGTGAAAGGCCCCTGAGAAAGAGTGAGGGAGTTGCAAAT
GTCAGTAGCCATCAAGATCTTCTTTAAGAATAGTTTCCACTAAAGAGATG
ATTGCTTTGGTTTCCAGCCTTCTTTGTTTTGTCTCCCCGCTGGGCCTTCT
ACCTTTAAAGGGCTTTGGCTCTGGGGGAATTGAGTTGGCTGGGGCTTGAT
GACTTCCAAGAGGACACAAGTGGAGATCTACTGCCTGCTCTTGGCTAACT
ACCTTCTTCAAAGATGAAGGGAAAGAAGGTGCTCAGGTCATTCTCCTGGA
AGGTCTGTGGGCAGGGAACCAGCATCTTCCTCAGCTTGTCCATGGCCACA
ACAACTGACGCGGCCTGCCTGAAGCCCTTGCTGTAGTGGTGGTCGGAGAT
TCGTAGCTGGATGCCGCCATCCAGAGGGCAGAGGTCCAGGTCCTGGAAGG
AGCACTGCGGAGAGAGCGAGGGAGGGAGCCTGGTGAGGTGGTCCTGCCAG
GAACCATGCTTTGACATCAGAGAGTAGAAAGCTCAGAGAGGAGGAAAGGG
CTTGAAAGAATCCCGAGCTTCTAAAGATCATCCCTCTCTGGGCCAGGCGT
GGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGAGGTGGATGAA
TCATTTAGGTCAGGACTTCAAAACCAGCCTGGCCAACATGGCGAAACCCC
TTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGGGTGCACCTGT
AATCCTAGCTATTCAGGAGACTGAGGAAGGAGAATCGCTTGAACTCAGGA
GGTGGAGGATGCAGTAAGCCAAGATTGTACCACTGCACTCCAGCCTGGGC
AACAGAGTGAGACTCTGTCTCATAAAACAAAACAAAACAAAACAAAACAA
AATAAAATAAAATAAAATAAAAAGATTATCCCTCTCTGAAGCTCAAGGAG
GTTAAGGGTGTACTCAAGGGCACACAGCAGGTTAGAGGCAGACTCAAGAT
TAGAATGTGGGCTTTCTGACACCTTACAGGCTATTCTTTTAGAATAAATC
CCATTTCTACTTTGTTCATCTTTTTTGTACATGCCCCACCTACACCATAC
ATGTATACCTTCTCTATATCTTTTTGTATCCCTAATGCTGTCACACTATG
ATTTGCTTTTTCATGCAGATGACCATAACATTTTCCATTCACCTATGCTC
ACTCAGCAAGTATTCAATTTTTCTACACTGTTCTTTTTTTTTCCTTTTTCA
TAACACTGTCTCATAGGCATTCTGCAAATCCTGTGAGAGTACTTTTTGTG
AAATGTTACCACTTTCCTCTTATTCAGAGAAGCTCCGTATTAAGGCTTCA
CTGAGGTTGCCTTAAGGCATGATAATGGTTCAAAGGCTTGAAAGACAGTT
AAAGAGACCTGTAAGTGCACAAAAGAAAGTTGAGCAGGAGAGAATTTCCT
GCCTGGAGCAGAGCCAAGCTGCTGGAAGAGGCAATGGGGGCAAAGGCCAG
GCAGACAAGCCAATGGGCTCCTCCCACAGCTGCAGCCAACAAGTTATGCC
AGTCTTAAAACTTCTAAAGAAATATGTTTTTAACAAGATTGAGGACTGGA
TTATGAGGCTAGGGGAGGCTATCACAAACTGGAATAAAATAAAGCCAGAG
AAAAGTGGCTGCCTTCCAACCTGCACAACTGACCTAGCTAGGCTGATGGC
TGGGCCACCTAGGAAGGCTACTGAGCATCATATAAAACAGAAGGGACAGC
AGGAATATAACATGGCTCTTTGTAAGGATGAGTCTGAAAAATGACCATTT
GCTGCCCAAATGCCCTTAGCTACAACTGAAAATATTTCAGAACTGGAGGT
TGCAGGATGCTGGAATCTCAGAGATCATCCAGCTCAGCCCTTTATTTTTC
AGATGAGGTCCAAAGCGGGTAAAATGACTTGTCAAGGTCAAACAGCAAGT
GAATGGTTTTCTTTCAAGTCTCAATTCATCTTTTTGTTTATATCATCTAT
GTCTTGTTGTTATAAGCTTCACCCCAGGTAGCAAAAAACTATTCTACTCA
AAAGGGGTAGACATATGTTAGTTCTCAAGATCATCTCTTGGTTTCAGAGT
TTAACTCAAGTGATTGGCATAGGCTGAATCCATCTCTTAAAAGGATAATC
AAATTTATGTTGAAGACTTGGTTGTCTTCCTACTATGAAATGGGAAACAT
TATCACTACTCCTCCCCTGTCACCACCAAGTGTGGCCACCACCACCAACG
TTAGTGAGTGACTGTGGTGATATGATGACCAAGTGGCCAGGTCAGCAAGT
GGTGCAGCCTGTGTCTCACTGGAAGAGGTTAAAGTCTTTCTAAAACAAAA
TACCATGGCATCAAAGTGGCCCAGAACTCCCTTCTTTGAGCTTTCCCTGT
GTTAGAGCCCTTCCTTGGGTTGGGAGTTAAACCCATAGTCTTACCTTCAT
CTGTTTAGGGCCATCAGCTTCAAAGAACAAGTCATCCTCATTGCCACTGT
AATAAAAACAGGGACATGTCTCAATTATGTCTTCTAAACAGGTTTATTTT
TCCTTCCCTGTGTACAAGACTTGACTGTTCATAAGAAACTGCAAACAGCC
TGCCTCTCAAAGCTGCCTGAAACACCTGGCAAGTTTCACAGTGATATGCG
CAGAACAGTCCAGAAGGCAGATTCTAGGCCTGGCAGGTGGGCACCCTGGG
TGCTCCCTGTTGGATCTTGAGGCCTAACCTCTAGCCCAGCAGAGTCAGCT
AAAATCTGAGCTCTCCCTCTCCCTCCAAGCCACACTTTGCAAAGGATTC
CTTGTATTGTGGGCTTGGAATCTTTTCTCCCCATTTGCCTCTGCAGGAAG

FIG. 3 (43 of 52)

CCCTTGCAACAACACA TGGATAGCCTCCAGGTCCCAAGGCTGGAGG A
CTTGTAATGGGAAAGTAGTCTTTAAATCAGATTTACTTGGCACCCTGTTT
GCCACTGAAAGAGGCAATTTAGGGGAAAAATCTGGTCTCCAAGCACAGAT
AACACTCTACTCTTGAAAGAGGAGACCTGCTCATGTTACTGGTCTCAGCG
TCTCCACTGACCTGTAATAAGCCATCATTTCACTGGCGAGCTCAGGTACT
TCTGCCATGGCTGCTTCAGACACCTGTGTAAAAAGGAGAAAATGAGTGAC
TTCCCCATGACGGCTACGTTCATGTGTGATTTCTCTCAGCATCCAGTGCA
TGGCAGTCATGCAAAGAAATGATCTCTGAGTAAATGAATGAATGTGTGAA
AGAGAAGTCCTTTGGGTCTAGAGAAAAGCATTTGCTAAACCAAACCCCAA
CTAGCAATGTATTGGCTAGGAGAGCTGGAGCAGAGGCTTTGACACTAACC
TTTAGGGTGTCAGCTGTTAGATAAGCAGTATCCATTCCCAGAATATTTCC
CGAGTCATAAGCATTATATTACACCTGGCATTTTTGCAAAAAGCTGAGAG
AGGGAGGCAGAGAGGGAAGGAGAGGGAGAGACAGAGAAAGAAAGAGAGAG
AGAGAGAGAATATGCATACACACAAAGAGGCAGAGAGACAGAGAGACTCC
CTTAGCACCTAGTTGTAAGGAAGATTAAAGTCATACTTGAGCAATGAAGA
TTGGCTGAAGAGAATCCCAGAGCAGCCTGTTGTGCCTTGTGCCTCGAAGA
GGTTTGGTATCTGCCAGTTTCTCCCTCGCTGTTTTTATAGCTTTCAAAAG
CAGAAGTAGGAGGCTGAGAAATTTCTCTGTTGAATACCTGATTTCACAAT
CAAGTTAAAGGAAAGGGGAAAAGAGTATTGGTGGAAGCTTCTTAGGGGAG
GGGACTAATAAACTGAGATAATTCTCTGGTTCATGGAAGGGCAAGGAGTA
GCAAACTATGACACATTTTGCAAATGTATCACCATGCAAATATGCATTGT
TTTCCTGACAATCGTTGTGCAGTTGATGTCCACATTAAAATACTGGATTT
TCCCACGTTAGAAGAATGTTTAAATTTAGTATATGTGGGACAAAGTGGAA
GACACACAGATTTATACATGCACATACTTTTCTTCATTCACTTCTTTGTA
CTTAAGTTTAGGAATCTTCCCACTTACAGATGGATAAATGGGTACAATGA
AGGGCCAATAGCCCTCCCTGTCTGTATTGAGGGTGTGGGTCTCTACCTTG
GGTGCTGTTCTCTGCCTCGGGAGCTCTCTGTCAATTGCAGGAGCCTCTGA
GGAGAAAATTGACCTTTCTTGGCTGGGGCAGAGAACATACGGTATGCAGG
GTTCAGGCTCCTGACGGAGTTGGGGCAACCCTGGAGATAAGCTCACACAA
CCCTGCAAGACCAGGTGCTGTTACCCTAGCCAATCTCATGGATGAACCAG
ATCAATGCCAGATGAGCTCTGCCTAAAATGATTTTTTGGTGAACTCTGAA
AAGTGGAATATTGTTTCTGTAAGAATATCCATCTGAGACTCTATCTCTTG
GTAATACCAAGAGTTATCAGTTTCTCTTTAACCGAGACACCAGCAAAGTG
CCTGCTCCAGGGTAATGCCCAGGGGAGCCCTCCATTTGTAGAATGAATGA
GAGTCCAGGTTATGAACAGTGCCTGGAGTGTAGGAACACCCTCCTTTGCC
TCTTTGACAGGTCTGCATCATAACACTTTTTTTTTTTTTTTGAGACAGAG
TCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGATCTCGGCCCCCT
GCAAGTTCCGCCTCCCGGGTTCACACCATTCTCCTGCCTCAGCCTCCCCA
GCAGCTGGGACTACAGGCACCTGCCGCCACGCCCGGCTAATTTTTTTGTAT
TTTTAGTAGAGACAGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCTC
CTGACCTTGTGATCTGCCCGCCTCGGCCTCCCAAAGTGTTGGGATTACAG
GCGTGAGCCACCGTGTCCAGCCTGTAACACTTCTTATAGCACTGAGTTGA
AACCTTGCTCCTCCTGGTTCCTCCAGGAAACTGAAATCTTTTTGAGCCAA
GTCTAGCACAGTGCCTGGCATGTACATTCAGGTGGTAGAGTTTGCTGCTT
GAATGGGTGAATGGGAATTTGACAGCATTTTTATTCAAATTAGTATGTGC
CAGGTATCGTGCTCGCTCTGCATTATCCAAGGGAGTGAGCCTCTGTGCAA
GTATTTGAGACACGAGGGAAATAGGTTCTACTGTGGGAAAAAGAGCATTT
CATGGACTTGCTCTCCAAGCAGCCTTCTGATTTTTAATTTGGCTCCCAGT
ATCTTGATATCAGGAGTCAGTCACAAGAACTCCATCTTTAGTAAGTTATA
TTTTCCACAGGAAATCTAAAAGCTGTTCAACATGTTAGTTTCCTGTGAAT
TTGATAAGCCATAATCCATTCCTAACACTGAGCCCTCCTGAAATTTGGTG
TCTGGTCCTGCAGATAGCTAAAAGCCCTGTCTGGGTGGCCTAGGGACTCC
TCTGTTTTGCCTCCACAGGATCCACTTTGCAAATTAACCACTGGTTCTCC
CGTTGTAGGAACTGCCACCTTCCTCAGAGCCTGTCTTTCTTCCTTCCTTC
CTTCCTTCCTCTTTCTTTTTCTTTCTCTCTCTCTTTCTTTCTTTTCTTTT
CTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT
TCCTTTCTTTCTCTTTCTCTCTTTCTCTCTTTCTCTTTCTTTCTTTCTCT
CTCCCTCCCTCCCTCTCTCTCTTTCTTTCTTTTTCTTTCTTTTCTCTTTT
CTTTCTCTCTTTCTTTCTCCCTCCCTCTCTCTCTTTTTCTTTGTCTCTCC
CTCCCTTCTCTCTCTCTTTCTCTTTCTCTCTCTCTCTCCTAGACAGGA

FIG. 3 (44 of 52)

46/118

TCTACCTTTATCCCCC..GGCTGGAGTGCAGTGGTACAATCATGCATTC..
TGCATGATCACAGCAGCCTCAAACCCTTCCTCAGAGTCTTTATGCGGCAA
CCAGCAGGGTCTGGAGGGTTGGTGGCTCTGTGAACTCTCCTGACAGAACA
CAGAGATGTCTTTGGTCTGTTGATGTGATTACAAGCTGAACGAAGGAAGA
TCAAAGCCAGTGACAGGAAGGGAGATATGCAAGGGACCCGAGCATCAGCT
CTGAGTTAGTCCATTCTGCTTCTGGGACTTGGGATACAGGTCAGAAACCT
TGAGCTTCTACTTCTCCATCTTCCAATTGTAGCATCCAGGACCTCAGAAT
CTGCCAGCTAAGAGGAGCCCTAATGATTGTCTGGTGGGATATGGTGGGAC
CACAGAGATGAAGACATGAATAGCTATTTGAATGTGAACAGCAGACGAAG
AAATCAAGGCTAGGAGGGTGGAAGTGACTCATCCAATAGCACAGTGTGGT
TGAAGCAGCACTAGTATCCAGGTTGCATGAGCCCCTGATGCTTTCGCTCG
AGGGAAATTTTGGAGCCATGGGGCAATGCCCCCTGACGTAACAGTCTCCA
CAGTTCTGCCATGTCTCATCCTGGCCCTGTAACCTGGACCCAAATCTGCT
ACCATCCCATCCATCTCAGGAAGTGAAACCTCTTATGTCAAATAGGTTGT
GCAACGTATGTATCAGATCCTGTCTTCCCAAGGAGACCGCTCAGGCCACA
GCACTTCCTTCCGATCCCCAATGAGCAGAAAATATCTCGCTATAAACATA
GTTGGCACTAAGGGAGGGAGTGGAAGAGTGATGATGATGTAGATGGTGAT
GTAGCCCCAAGGAAGTGGAACAAGCAGAGATGGGGAGCTGGAAATGCCAG
GATGCTCCAGCTTTTGGGGAATTATTCAGCTCTTGAGTCACTAAAGCCTT
TCTCAGCTGCAAGTTCCTCTTTACCCTGTCAGGTCATTCTTCCAAGACAG
GAGACTGACATTTATTCAAAGCAGCAAGTGCCCTGATACCATCTTGTGTC
TAATCATGGGCTTCGCAGCCAGTTATCAAGGTTGATCTCATCTCATTGGT
CTTCAATCATTTTGAACAAGAAGACAAGCAAAATAATCATGGGTTAGTTC
TTATATTATTGTGTGTACATGCAGTGATGTCTGTTCTTTGTAGTGAGCTG
TTCCTTCCTTGTTCACCCTCTTGCTTAGAACAGAACTAAGCAATCTGCCC
CCAACATTTTCCCCAATTTCCCATCTCATTCTTGGCACTGGCTTCCTAAT
ATTTGTTCTTATGAGTCATTTTCTTGTATCATTTCCATGAGTCCCTCTGG
GATCTTAAAGTATGAAAAATGTTGTGTGTACCCACACCTGTCTTTGTGGA
TATTTCTCTCCTTTCCCTTCTGCTTCTGGGATTATTTGGGAATGGGCACT
ATGATTTTTATCATATCGCTTCCACTTCCTTTATGGCATCATCTCCAATG
GGCTTCTTCTCCCTCTTGGATCCAGGTTCTCAGATTGGGGACATGCAGAG
TCCAAGGAACATTCCATTCTCCTCCCTGGTCTAGAACAAGGAGGGCTTAG
ATATATGAGCAGGTGGCTGGGGCTGGCGAGCTATGTAGTCTCCAATGGCT
TTTCCCTGATGTCGGAGTTGTTATGTCAGTTCTGGGAGACCAATAAGACC
TTGTCCTTCCTTTGGATCCATCTCAGAAAAAGCCCCTGGGTGGGTAAGATGG
ATGGCAGGGCTCTCCTACTCTATGTCTTTTCTCACACCTAGTGGGTATAA
GAGAGGGGACCACAAACAGAGGGGGCTCTGGTACCACTTATCCAGGGTCT
GGAAACATTTTCTGTAAAGGGCCAGATAATAAATGTTTCAGGTACAACTA
CTCAACCTTGCATCATTTCAGAAAAGCAGTCAGATAATACATAAATGAAT
GGGTGTGGCTGGACTTGTCCTGCGGTCCCCTGTCTTATATCATTGTATTA
TATCATTTTTTCTTACATACAAATTTAGAAGCAATACTTAAAAAAAAAAA
GCCGTCCTTTATTGAGCACCTACTAAGTGCCAGGTACCTTTTTTTCCCTC
ATTATCTTATTAACTCTTCATAATAACCTTTAAAGTAGATAATATTGAAC
CATTTGACCTATGCAGAAACTGAGGTTGAGACAATAAATTATTTAAGACC
GCACAAACAGTAAATGCTGGAACTACGACTCAAATATGGGTTAACTGAAC
CAAAACCAGATCTTTATTTCTCACTTTTAATTGTTACATATGTTTATTGC
CTCATCTCCTGTCCACATGGTGCCCATCGGCAGACTCCTTTCTCATTCTC
AGTGATTGAGTGACATTCTAAACTACATTGGCCTGGCAGATTCACCTCTG
TCCCCTAAATGTTTCCACATTGTCCTTTTAGGATTGAGATCCTCTCTGTT
CCCTTGTCTTCCCTCCTTTCTTCTTCTGGCGGTGACGTGCTGTGTGAATT
TGTTTCTTTCTCCTCTCAGGGTAGTACTGGACTTTCCAAATCAGGGTTT
TTAGTGATCTCTCTTCCCTTTTCTGAGTTTCTTCCTTATTCCCATTCACT
TTCTCATCTATAAGTGGCAGCTTTGTTGCTGGAGGATTTCCTTTGTCCTT
TTATTCTTCTTTAAGACTTTGTCATAACTGTCAAAAGCAATCCCTTGAAG
GTATCTGTCCTTGGAATTGTGTGCTTATGATGCTGAAAAATACTCTCTTC
CTAAAGCTATTATAAATGCT

>Contig42

GGCTAGCTGCAACTCTTGAATACAAACACATTCAGACATGCACACACTTT
CTGGCTCCCAAAAAGAAAAAAAAAAATCAATTTATAATAATTCTGATCCT
TTGCTTATTTCCACAAACTCCATGAAAATTGTACATTGTCCAAGCAACAT

FIG. 3 (45 of 52)

TTCTTAATATTCTCTT..TCTCTCATATCCATTTTCCTTACTGCTGTC.C
CACCTATCTCTTCCAAACTCCCTGTTAAAATCCCTGCCCCAGCGAACTTT
TATTCAATTTTGTGGAATGGAGGCTGCACTGATTTAAATTAAAAAAAAAA
AAAAAATCCCTACTCCATGTCCCAGATCCCTAGTTGTTTTTTGTTTTTTG
TTTTCCTGAGACAGGGTCTTGTGTCTTCCATGCTGGAGTGCAGTGGCATG
ATCATGGCTCACTGCAGCCTCAACCTCCTGGGCTCAAGTAATTCTCTTGC
CTCAGCCTCCCCAGTAGCTGGGAGTTCAGGTATGTGCTACCATGCCTAGC
TAATTTTTTTCTTTTATTTTGTAGAGACACGGTCTTGCCAGGTTGCCCAG
GCTGGTCTAGAACCCCTGGGCGGACGTGATCCGCCTGCCTCGGCCTCCCA
AAGTGCTGGGATTACAGGCGTGAGCCACTGCTCCCGGCCTTGGGTGCAAA
TTTGAGCTTTCTCACTTATTAGTGTAAGACATACAGCTAATTTCTAAATC
TTCCAAACCTCAGATTTTTCATCCATGAAGTGAGGATTATTATAGAGCTC
ACTAATAACATGGCTTCAAAAATATATAATGCCAAAATTGAGATCAAAAT
AATAAATCTATATTACATGGGAGATCTTAATGTACCTCTTATATTATTGA
TAGACTAAGATGATCAAAAAAATAGAAAGAGAGCAGTAAGGAGAGCAAGC
ATTTAATCAATAGGACCAATACATTTTAATCAATAGGATCCTCAGGAATA
TATACAGAATACCAAACCTAACAACTGCAGAAAACATGCCAAACATTTAG
GTACAGACATTGTTGGAAAATGCAATCTTGAAACGAGTGGACTGACATTC
AGAAGATATTAATAAGAGCACTAATGATGGGGATTGCAACCATGTCTTTA
CTGACTTCCAGAAGCTTCTTACAGTAAACATGAAATCACATAATTTCTTC
CACTTTCCTACTGTTTCTTGTTCTGGGCTCTGTCCTGCTTACTGTCTAAT
ATCTTGGCCCCTTAAAAGTTGCTAATCTTCCAAACCTCATTCCTGTGACT
GGGCCGCTGGTCCTTGTTCATGGGCCTTGAAAATACTGACTGTACACTTA
TCTGGAGCATCCAGTGCCTACCACCTGACCCAGATTCCTCATTGCGCTCC
TCCCTCCTCCACCTATTGGAATTTGCTCATACCCGTGTGAGACCCCTCCC
TTTCCCCCCATCTGAATTTTTATCAAGACAACGCACTGCCATACTCCCTC
GTACCCTGCTCTGGGCATCAGACTGAATGTTTGTTTCCATTGAGGATCTG
CAGCTGCATCAGTTTCCCCAGCACCGTCCAACCCCTTGAGCATGGCTAGT
CCTAAAGCAGAGAATTAGCCTTTCTATCCCTGCTGCTATACATGCTGGGA
CAAATAATAAGAAATGACAGCATTTTATGATAATGCAGGCTGCAGGAGGC
AGGAGGCAGGAATCAAATTCGTGCTTATCAAATAGTGCTCCAATTCTTTG
AATATTGGACTATAGAATATGTCATGGATCTATGCTCAGGTGGGTTCCCT
ATTACTCACTCCACTGAGGCCAGGTTGTGGGATTAGCTGTCCAAGAGGGA
GTTTCAGTCTCACAGCATAGGGTCATTCTGAGAATTACTGGCCCACACTT
GTGTGGAGACCTCCAGAGAACAGAATCTGGGTTGGTGCCATGTACTTCCA
GGAGGAGAGAAGTGGCAGGATGCCCAGCCCCACAATCAGAGGGGAAGGGG
CAGAGCCACATGTATGAAGATCCTCTCCCCAGTACGTGCCAATCACAGGG
CTTCCTAGCTTTTGGGCCAAGGAAACAATGTGGGAAGCAAAAAAAGGACAA
TTTTCTCCTCCCTTTGCATGAAGACTGAGCAGTTTTACCAGATTCCCAGG
GAAACACCCCTTCCACTCTGGGTTGAATGTGAGTGAGAGACATTCAGCTGG
AACACTAGAAAAACTATTTCCTGAGCCACTCACCTTTAGCCCTAGAAAGT
GTTGGATTTGTCCTTCATCTTTGCCACAGTAGAGACTGCTGATAGCATCA
GAACTTGGGCTCTGGAATTAGACAGATATGGGTACAAATCTGAGCTCTCT
CACTTATTAGTGTGGGATGTAGAGCAACTTTTAAAATCCTTCCAAACCTC
AGACTTCTCATGCATGATGTGAGGATTGTAATAGGGCCCACCTAATAGGG
GTTTTTGAGAATTAAAAAAGTTATTCAATGAACAGCATTTAGCAAGATGC
CTGACCATTGAGAAAATAACAAATTGTTTATTATTATTGTTATTATTAAA
CATCTTTCCTGCACCTTCTGACTGGGGGCATCGTATCATCAGAAATACTT
AGGATGGGATGGATTCCTGCATGGGCTGAGTCAAGGGTGCAATAATGGAG
GAGTGAAGAAGGAAGAAATGGAGGCAGAAATCCCCAGGAGCCCAGCATGG
TACAAGGCTGAGCTAGTGCTGCAGAGCCTCCTTGGAACAGCCACAGAGCT
TGCATCTGGCCCTGGGAGGAACCTCTTCTAGCTGGCAGGACCAGCCACAA
CAGTGGCCAGGGGATTTCCCAGGGCGTGGGCTCCTAGGAGTTCATTTGGA
CCAAGCCTGCCTGGAGAGGGGTTATAACAGGGATCCTTCCCTACTGGCAG
GTGATTTACCCCTCGGTGAGAAGCTCAGGCATTTGTTTGATGGAAGGTGG
AAGGCCCTGTGCTGGGCCAGTGACTATCAGGGATGGGCGGTGGCTGGAA
AATAGCAAATAAGACAATATGATAACACAGTTAACCACCACACTATGTGA
AGCTACAATATGGGTATCTGTAATAGACAATTCCAATGTAGAGAATAATT
CTAAGGTGTCATTCTCCCCGCCAATGCCATAAGCACACGGCCTCTGCCTG
GGTTTCTCACTGTGGAATGTCCTCCTGGTCTCCTCATGCCCAGAGAGTGG

FIG. 3 (46 of 52)

GAAGTACTCCTACTTT[illegible]CACCGGCTTTCCTGTCATCTCCCTGCAGCC[illegible]
CCTCAGCCCCCTCTGCACAGGGAGGTTTCCTCCCTGCTGCTGCAGTGCTT
TGTACTTGTTAGTGGTACCTGCACACAGGTATTGGTGTCCTTGTCTCACC
ACCCTACATCACTGTAAGCTCCCCAGGAGCAGGCTTCCTGTTTGACTCAC
CTGTGATCCTCCACCTCCCACCCTGTAGTGCCTCAAGCATTGAGGACAAT
CACTGGCTGCCCCTTAACCCAGAAATGCTGCCGAGACAGGAGGCCATGGC
CCAAGTTCCTGGAATGGGGTATTACTATGTCAGCACAAAGGCCTTTGCAC
AAATGAAGGCTTTAAAAATGCAGTCCTAGTCAGGTGGAGGAGGGCTTATA
GGATTCCCAGGAATCTGGATCATTCTCTTGAGAGCTTTCCCTTGTCTCTG
TTAAAACTCACATCCTACGGCCCAAATAACAACAAAAAATGGATGTAAAT
TCTTGAAATAACTTGTGGATGGGGGAACAAGGCCCACCCCCCAGATCTGC
CAGAAGCTTCAGGTGAGGGTCCCAAATGCCAAAAAGTCTGGTATCAGAGA
GGATGGCCAGTGACCTGGGGACACATGCCCTTTGCTGTGTCACTCAAGGA
GCAGCAGCCTCGGCCCCGCACAGTGACCAGGACCCTGGCTTCCCACGCTG
GGCAGGAGCTGGTGTCTGATGAAGGGAATGCCTGGCAGCACGTGCTGTCT
GTCTCCTCGTGTCAGCTTACCTGGCTTTGCTGCGAAGAGGCCACTCGCAT
TTCTCAATTTTTTATATTTTTTTAATTTTTTAAATTTTTTATTTTATTTT
TATTTTTATTTATTTATTTATTTTTAATTTTTTTTTAATTTTTTAAATTA
TGCTTTAAGTTTTAGGGTACATGTGCACATTGTGCAGGTTAGTTACATAC
GCATACATGCGCCATGCTGGTGCGCTGCACCCACTAACTCGTCATCTAGC
ATTAGGTATATCTCCCAGTGCTATCCCTCCCCCCTCCCCCCACCCCACAA
CAGTCCCCAGAATGTGATGTTCCCCTTCCTGTGTCCATGTGATCTCATTG
AATTTCTTTAAAGGTGGAATCTCTCAGTGGGGTCTAATCTGTTCAGAAAT
ATCAAAAGAGTATCCTTGGGAATGACTGGAATTCCAGAGTCATCTGGTAA
TCCTCATAAAACAACTCCTGGATGTCTCTCAGCACATCTCCCACCTTGAA
CGCAGGAGGCTGGTTCAAATGGAGGAGCATCGCTCTACTGCACTTTTTTT
TTTTTTTGGCCTAAAGTGCAAAAGGGGATACGTTTCATGTAAATAAATCA
ACTGCAAATCGCTAGTTATGCTGAGCCCTGTCCCGTGCTGTGGACACAAA
GGAACCAAAGGCTTTTCTCCCCGCCCAACACACACATAACACACACACAA
AATCATAAAAACATACATACCCCCAACACATAACAACACACAACACACAC
ACAAAATATATACACACAACACACACCAAACATGCCCACAAACCTGTGTC
CAGAGATAGATCCTACTGGTGGGTTTGTGGTCTCGCTGACTTCAAGAATG
AAGCCGTGGACCTTCGCAGTGAGTGTTACAGCTCTTAAAGATGGCATGGA
TCCAAAGAGTGAGCAGTAGCAACGTTTACTGTGAAGAGCAAAAGGACAAA
GCTTCCACAACCCAGAAGGGGACCCCAGCAGGGTTGCTGGTTGGGGTGGC
CAGCTTTTACTTCCTTTTGGCCCCTCCCATGTTCTGTTTCCATCCTATCA
GAGTGCCCCTTTTTTCAATCCTCCCTGTGATTGGCTACTTTTAGAATCCTG
CTGATTGGTGCATTTTACAGAGTGCTGATTGGTGCGTTTTACAATCCCCT
TGTAAGACAGAAAAGTTCCTGATTGGTGTGTTTTACAATCCTCTTGTAAG
ACAGAAAAGTTCCCCAAGTCCCCACTGGACCCAGGAAGTCCACGTGGCCT
CACCTTTCAACTCCATAATGGCATGAAAATACATATGTTGTACAAAACAT
ACATACACAAAGTATACATGCATCTCCCCAAATATACACATACCACAGAA
ACATACACACAGGAACTCAGCTACCTGTCAAAAGTCTGCATGGTGATTGC
CTCTGCAGTGAGTAGTTAGAAAAGTGAATTTGTTTTTCAATAAATTGGAG
TCCTTAAAAATCGTTGTAAGATAGAAAATTTTTAAAAGTATATAAAATAA
AATATGTATGTCCTTTGGTCTAGCATTTACACATGTAGGAATTTATCCTA
GTGGAGTAATCAATGATATATGCAAAGATTTGGACAAGCATATTAAGCAC
AGAATTATGTATGCATATGTGTGTGTATATATATATATATCTCATACATA
TAATAATGTAAAAGTGAAAATAACTCAGATGTTCAAAATTGAGGATTAGT
TAGACTATGATCTGTCCATATGTGACATACAAGTTAGCTGCCCCTTATTC
TCTCGAGCTTCAACCTCCTATAAACAGTGTCCCTTGTATATCAGTATTGG
TACAGATAATCGAACTTATTGAGGTTTTACATGGGGCAATAAAGGCAAGA
GTTTATGAATACTCCATACTACACTAGGTAGCACCCCCTATTAAAGACAA
ACTCTTCTCTCTCATTTCCCTTCCTTTCCGGAACCACTTGGTTGAATCTC
TACAAGTCTCTATTGCAACTGCCTCAACATGGCACCCTCCCTGCATCTCC
ATCTTCCCTGTCCTGAGAGCAATGGCCTGCTGCCCCCACACTCACATCCT
CATTCATTCCAGAAGTGAGCACCACAGAAGTGCCTACAGTTACCCCAACC
ACCTTCTTAGAAGATAAGTTAGTGTTTGTTTTGACTTTTTAAAATTTTTA
CTTCCTCTTTTCCTTCACAATCTCATCCCATCCCAAGAGGTTTATCAAGA
AGTTCTCTAAAGATATGTGTCTCCTTATGGAATTTAACAGAAATCAGGGA

FIG. 3 (47 of 52)

...GTAT..C.AGCCATT GGGAATAACATTTTTCCAGGTCTTTAGAC
ATAATGGAATACCTTGCAGTAATTAGATACACTATTGTAGAAAAGTATTG
ATGAAATGGAACGATGTTTGAGATATCATATTGAGTAGAAAAGGCAAGAT
ACATTAAGTAGGAAATGTATCTTACAAAATAATTTGTCAGACACACTCCT
ATATTTGTATGTTATATAAATGCGTATGTGAAGAAAGGCTAGAGGATGAG
ACCACAGTCTTCGGTGAAGTTTAAGAGATGATGCTGCAGCATGCTCAGAA
AGGCTTGGTATAGTTTTTTCCAGTAATTAAGGACTGATCTTAGGTAAATT
GTCCATCCTCTCTAAACTGCACCACCTTTTGTCTGTAAAACAGGAAGGAT
GGTATTTACCCCCAGGGTCATCAAAGGATTTGGTTGGAGAAAAATAAATA
AATGGGCTGAGCCCAGACCTGGCACAGTGAGAGCACAGTGGTTGACTATT
GTGCTGGCCTGTTGTTCCTGTGTTATTGACATGCTGCTGGTGGTGGTCCA
GAAGCTATTACCTTAATTGGTTATGTGGATTTCCCCTCATACTGAGCAGC
TGTGTGTGGTGTTGTAAAACATAGCCATACACAGTAACTGACAAGGGCAA
ATGTGATGGAAAAAATGCAAGGAAGTGCAGATAAATAGCTAATGGGCTGTA
GAAGGAAGCTAGTCCTTGGAGGGCTTGATCAAGGAAGGTCCTTTTGCATG
TCACCTTTGAAGAAGAGGGGACATAGAAGAGGTATAGTGCATCCCGGAGT
GTACCTGGAAGGGAACATGAAAAGAGGACATTTTTCTCTGGGACATGGGG
ACTCCACTTGCATGAACTCTGGAATTGGGGCAAAGAACCATCATGAGAAC
AAGGGCTTCCTTGAACCTCCCAGGCTCATTGGCTGATCTAAACCCTGTGT
CCCCTCTTTCCTTCACTCTCCTCTGTTTTCTATACCTGTATTATTGGACT
GGACTGGAAGCCACCTGATCTATCACAAGTACCTTGAAATGTGTTGAATA
GGTGTGGCACAGTCCTTAGCAGAGTGGCACTACCCCCACAGGAATTTGTT
TATACCTTTGGCATGGAAAATAGCAGGAAATGAGTGATCACTGATAACTG
AGGATGCTATTTATTATTGGCCAAAGGAATACTTGTGTTGTATTTGCATA
ACCACTCACAAACTGTTGATTACAAATGAGTACCAGACCTAGCTCCTTCA
AGTAAAGGATCTTGAGAACTGAAGGCAAACAGAGCTCCAGGAGTCCAAGA
CAGAGCCACAGACCACGAGGATCCCTGGCCCAGGTAGGTGGTCCTCCTGC
ACTGGCTTTCAAGGCCAACAGGATGGATGGGGAAGTAGAGTAGCATCTGG
CCATCTAGACCCTTGCTTTTTATCCCCACTGGAAGCACATCTGAATTTCT
AAATATGATCTCTGAGACCTGCCCAGAACACCTTGCTCTCAGCCCCAGTA
GCAGCCTGCTCTCTCCCAGGAGGGCTTCCACTAACAAGTAGGGCATTGCT
GGAGGGCCAGGCAGACACTAGCTTAGGAAATCCACCAACCCTGGAAATGC
TAGTCCCTTCTCTGAAGGCTCAGAAGACTGACTTTAGAGTCTAGAAAATA
TTGGTCCTTGGGAACAGATTTTGAGTGCAAAGAGATGGACTTCAGATGGC
CAGATGCACTGCTTCTTTAGGGAATTCTGTGAAAGCTCCCTGCATTTATC
TTAATACAGGCAGCAGATTTCATGAGTACCCCCGAGGGATGGCCCCAGGT
CCTCCAGCCTGTGAGCATCCTTCTGTCCTTCAGCAGCACCACAGTATCTT
TATATGTCTTTGGATACCTACGTTTCTGCCAGACATCTCTTGCTCTGATG
TTCTGGCTGCCAAATTCTCTGTCAAGCGCCTCCAATTTTTTGTGTCCTTT
GATTTACCCCAACATGACAAAGGCAGTTGTGCTTCATGTATTCAGGGATA
CTGCCAAACCACAAACAGGTTAAAATCAAATAGCAGATATCCCTGTTCCT
AAAGACCCATCAGCTCTACCCACCTGCTCCTGCTCACCGTCCTTATTGTT
GAGTCCTGAAGCCCTTCCTTGTCATTTTTATTTTTTGCATGAACAATTTA
GTTCCCTTTGTCTCACTCCTAAACCTTTCTCAAAGGATTGGATTTGTACA
CAAACTGCCTATCTCTGCAATCTTAGAAGTGATATGATTCTGAACAAATC
ACTTAACTTTTGATTTTTTATTGGTAAGATGGGAATACCAATTTTTGCTC
CACTTCTGTCCTATGTTGGCCTGGGCTGATGTTGAAAGCTCTCGGTCAAC
TGAGATAGGGTGTGCAGAATTTATATATATAAATATATCTCCTCCAACCC
CTCCCAATGAAGCAAGTCACGTGAGTCAATCCTACCCTAAGATATTAGGG
ATTGAGCCTCCTGGGACATTTGGTGGCTTAGGTTTTCATGAAAAGAGGTT
GCAGAGCAACTGCTTTTTGTTAGGCAAAGATTAGGCTACTGCAGAGACTC
AGCAAACTTCTATAGAAGGTGTCAGATGGTAAGTATTTTAGGCTTTGCTT
GCCAGATGATCTCTCAACTAGTTAACCATGCTATTGTAGCCTCGAAGCAG
CCAGAGACGATCTGTAAACAAGAGCATGTAGTGTTGGCATAAATATAGTA
CCGCG
>Contig43
GCAATAAGTCTATTTACTGTAAAGTTAATCAAATTTACATTTCAGAACAC
TTAATCTGCAAGAGTCCTTTCCAAGACCCTATACCTAATTTTGTGTTTAC
AATTTTATATTTGTTTTCTTAAAGAAGACCACCAATATAAACTATATCCA
GCCTTCATGATAAGTACATAAGAAACTATGCAAATAAGGGGGAAAAAAAA

FIG. 3 (48 of 52)

CAAAGAAAAATACCTAC TTACTAATGGTTCACTTCTGAATAGCACAT...
TCATAATGATACAAGCACTCATTACTAGTCTAGGAAAATGAAGATATAAT
TGCATTAGGAAGATCAAGAGGTAGGAAATGTGGATGTGTGTGGTATAGAC
TAGGGCAGGACAAAGAACCTAAATCCTCATTTTCTAAAGATAATTGTTAA
TACGTAAAACTCAAAATTCAAGAAGTAACAGTAAAAGCGGTCATTAAGAA
ACAAGCACTAAACACCAGATAGGAAGCGAGAGATGGGGGAAGAGGGCAAG
AATCTGATTATTTTTTGCAACAAATTTTGTAAAACCATTTGACTGTTTAC
ATGTAGAACTTGGATCTTTTTTAAAAAACACAAAATAATAATACTATTAT
TTTTTAACTGGATTTTTGAAAAAGAAGATAAAAGTCTCATTTTAGTAATT
AAAACTCATTCCAGGTTAGTCCACTCAAAACTTATATTCGAAAATTAAAA
CTTTGGGAGGCTGAGGCAGGCAGATCACCTGAGGTTGGGAGTTCGAGACC
AGCCTGACCAACACGGAGAAACCCCGTCTCTACTAAAAATACAAAATTAG
CTGGGCGTTGTGCATGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAG
GAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCACA
CCATTGCACTCCAGCCTGGGCAACAAGAGTGAAACTCCATCTCAAAAAAA
AAAAAAAAAAAAAATTAAAACCTCTGGAAGTTGAGTTTGCAGATATTCAT
TATGCTCATTTTTAACTTGTATGTTTGGAAAATGTCATGATGAGAATTGA
GGTTGGGGGATGAGAAAAAAAGAAAAACATCAACCCCACAGCCCATTCAA
TTTTCAGCCCGACCCACAGCTCCGGGGAAGGGCAGCAGGTCCATCCTTCA
CTCTTTCTTCACCTCTTTCCCCTCCTTCTGGCTCTTCCACCTCTAAGTTG
GAGCCCAAGAAGAGGCACTGGGAAATGGAAAAGTCTTTTGTACGTGGTAC
TTGCCGGGGAAGCTGCCATGAAGACCTGGCCCCACGGTGGGGAGGGAATG
CCCAGCTGAGGCCTCGTGCCCATGCTAGGATAGACTCGTCCAGACATGTC
AGGTGGTCTGACAGGGCAAGCAGCAGGAAGTCATGTATGAGTATGAACTG
ATCTGTATGCAAGGGCGGGGAGAACACGCGGAGGAATGGGGCGTGAGAAA
ACAGCACAGTACGTTTCTTTAGCAGCTGTCTCTGCTCAGCCATGGGAGTC
ACCAGAGAAAGAGGCTTGGAGGCGTTATTTTCACTGTGAGATGTGAGTGT
AAAAAAGTGCCCAAGACACAGTGAGTACCAGGGAGATGCCCTCTTTCCCT
ACCCGAATGCAGAATGGCCACAGGCCTTAAAACACACACATGGTTCCTCA
GAGGAGAGAGGCCTCCACAGTGGACACCCGCATTCTCCCCTGGTCAGCAG
CAGCAGGGCGAGTGCTGGGCCATCATGAAGCTTCACAGGCAATGAGCTCT
CAGCAATAACAGGAACAGTGCCTGGGGGACTGTAGCTGCAAGACCGATTT
TCATGTAAGATGGCCTCTGAGGACTCCGAGATACACCAGGCTGAGACTAG
CTGGCAGCTCCAAGTTCTTGGTCAGAAGAGAACAGGAACTAGGGAAATTG
GAATTACTGTTACTACAATTCCTTTACATCCGCACAACCATGAGGTCCAG
AGAGTCTCTCTTATTTTTTTTTTAAAGACAGGGTCTCACTCTGTCGCCCA
GCCTAGAGTGCACTGGTGTGATCATGGTTCAGTACAGTCTTCACCTCCCA
GGCTCAAGTGACCCTCCTGCCTCAGCCTCTCAAGTGGCTGGGACAGCAGT
TGCATGCTACCAGGCCTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTT
TCGGTAGAGACTGGGTCTCTCTGTATTGCCCAGGCTAGTCTCGAACTCCT
GGGCTCAAGTGATCCTCTGGCCTCAGCCTCCCAAAGTGTTGGAATTACAG
GCATGAGACACTGCACCCAGCCAGTATAGTCTTTTAACAGCTTTATTGAG
GTACGGCTAACATTGAAAAAACTACACAAATGTAAAGTATGCAATTTGAT
AATTTTGACAAATGTACACACCAGTGAAACTATCACTACAGTCAAAATAA
TGAACATATCCATCACTCCCAATTTCCTCACGCCCCTTGGTAACCCCTCT
CTCCCAACTCCCTGCCCCCTAACATCAGACAACTACTGATGCATTCTGTC
TCCATAGGCTCATTTACATTTTCTAGAATTTTACATAAATAAAATGACAG
AGTATATACTCCTTCATGTATGGCTTCTTTCAGCCCAATTATGTCAAGAT
TCATGCTTATGGCTGTGCGTATCCTTAGCCCATCTCTTTGTCTTGCTGAG
TAGGATACCATTGCATAGACAGACCACAGCTTGCTCATCCATTCACTCTT
GACAACGTTGAATTGTCTCTGTTTTTTGCAATGACAAATAAGGTTGCTAT
GTACATTCCTGTATAGACATTTGTAAAAGCACAGCATTTCATTTCTCTTG
GGTAAAGACCTAAAAGTGGAAAGGCTGAGTCATATGGTAAATATATATGT
CTAACTTTTTAAGAAACTGTCAAACTGTTACCCAAAGGGATTGTACAATT
TTACATCCCCACCAGCAGTGTATGAAAATTCCCGTACTTCCACATCCTCA
CCAATATATGGTGTGGTCAATCTTTTTAATTTTGGACATGNTAATGAGTG
CAAAATGAGGCCCAGAGTGTCTGAAGTTACATTTGTATCCTTTTTGGCAT
CCAAAACAGGTGTCAAGCATAGAAAAAACACTTGTTCCTTGAATGGTCAG
TCATTTACAAGTGGAATTCATTACAAACCGGTAGTTCTACTGGGTTAAAC
TATGCCTTACTGTCAACAGGCACATACACATACAGACAGACAGGAAGGCA

FIG. 3 (49 of 52)

CAGAGACAAGGCAGAGC...TGATAAGAAGGTGACCTGGGCTCTAGCTCT
GCCTATCACCTAGTAAAATATTAGTTAAGTAGCCATGAGTAACTCACTTA
ACTTACCACAGGCTCCATTTTCTTATCTGTAAAATAGGAACATTGAAACA
GCTAATCCCCAAGGTTTGTGGATAATCAGAATTACAAAGATCAATGACAT
TTCTATGAGAGAAACATATTTCCAAGTATTTGATGGAGTACATCAGACAC
AAAGGAAAGGAAACTGAATATTTTTGAGGTTTTTTTTTTTTACCAAGAAA
TTCACATTTTGTTAAATTTTCAGAACTACCTCCTGAGGAAAGTGTAGCTG
CACCCATTTAGAATGATAGAAAACATCAATCTGTCTGATTCCAAAGCCAA
GTTCTTGCTACAACGAGAAATGAAACAACTGGATCCCTACAGATGCAGAG
ACCTGGGCCCCACAAATGTGAATTCTGTTCCCCTACCGAATAGAGTTACA
GTTCCATAATACAGTACTCCCTCACTTTTCCACAGTCTCACATTCCACAG
TTTCAGTTACCCACAGTCAACTGCAATCCAAAAATATTAATGAAAAATTC
CAAAAATAAACAATTCAGAAGTTTTAAATTGTGCTCCATTCTGAGTAGCG
TGATAAAATCTTGTGCCACCATCCCACCTGTCCAGCTTATCGTTAGTCAT
TGACATCGTCTGCTCCTGACATCCAACCATTGACATCATCATGACTCTAT
GATCCAGGATCACCGAAGCAGATGACCCTCCTTCTGACATATCATCAGGC
CAATATCAGCCTAAACACTGCATCACTATGCCCACATCAGTCACCTCACT
TCATCTCATCAAGGAGGCAATGGATCACCTCACATCATCACAAGAAGAAG
AGTGGGTATAGAACAATAAGATAATTTTGGGGCAGGCATGGTGGCTCACG
CTTGTAATCCCAATACTTTGGGAGGCCAAGGCAGGAGGATCCCTTGGGCC
CAGGCATTCAAAACCAGCCTGGGAAACATAGTGAGACCTCCTCTCTCTGC
AAAAAAAAATAAACAAAATTATCCAGATACAGTGGTGCATGCCTGTGGTC
CCAGCTACTCAGGAGGCTAAAGTGGGAGGATCACTTGGTCCCAGGAGGTC
GAGGCAGCAGTAAGCTGTGATCGTGCCACTGCACTCCAGCCTGGGCAATA
AAGTGAGACCCTGTCTCAAAAAAAAAAGGTAATTTTGAGAAAGAGACCAC
ATTCATACAACTTTATTATAGTATATTGTTAGAATTGTTCTATTTCATT
ACTTATTGTTGTTAATTTCTTTCTTTGCCTAATTTTTTTTTTTTTTTTTG
AGTCGGAGTTTCACTCTTGTTGCCCAGGCTGTAGTGCAATGAGACGATCT
CAGCTCACCGCAAATCCCGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCA
GCCTCCCGAGTAGCTGGGATTACAGGCGCCTGCCACCATGCCCAGCTAAT
TTTGTATTTTTAGTAGAGGCGGGGTTTCTCCATGTTGGTCAGGCTGGTCT
CGAACTCCTGACCTCAGGTGAGGCCTCAGCCTCCTAAAGTGCTGGGATTA
CAGGCTTGAGCCACTGCGCCTGGCCTCTTTGCCTAATTTATAAATTAAAC
ATTGTCACAGGCATGTATTAATTTATAGGAAAATCATAGACATATAGAGT
TGGGTACTATCCACAGTTTCAGGCATTCACTGAGGGGCTTGGAACACGCC
CTCCTCAGATGAGGGGGGACTACTGTCATCTCCTCAATCATTCTTGATTC
AATCCTCAACACAAATGGTTTGGCCAGGTCTTGCCTCTGGAGACAAAATT
GCTAAGGATTTAGAGGGGAAAAAATGTAGTTCACTGGGAAAGTCACCTCT
GCTCCACTGGACAGCAACTTAAAACCCAGGCCATGACAAGTAGAAAGGCC
ACCCCCACTCTCCTTCACACCTGGAGTATTCAGGAGTCAATCATATTTCA
GGACCACCAGGAGCAAACTGGGAAAAACTGAGCTGCCTTGAGGAAAGCAA
TCAGCTCCACAAGGGGCTTAAGAAACAAGCTCTGGGAGGAGTGGTTGGAG
AAGAGTTGGGGACACATCAGAAATGCCATCAAATTTCTAAGGGCTACCTC
GTGGTGTCAGACCTGTGCATCTTCAAGGACATAAACAGATGGGATAAGCA
GATGAGATTCACAGAGGACATCAAAATATTGGCTCCCCAGAAGGGAGAAC
ATTCTAGTAACAGAGCTGCCCAGCTGCAGAGTGGACTGTTTCACAAAGCA
ACAGGTGCCCTGCCTCTTGAATCACCATCTTCACAGGAATGCAGTAGAAG
GGACTTAACTCCTGCCCTGAAGAAAAGGTTAGGCTAGGGAAACAGCTCCA
AAATTTTTAAAAGGAAGCAACATAGGCATCTACTGGGAGTTTTCTAAAG
CCTTTGTTTAATGAAACTAAAGAGCTGGGACAGGAAATGCCAAATTAAAT
TAATAGAGCCTTGCTTTAAGACAATGCAAGTGGATGGTAATGAAGGAATG
AGTCTTAGGCCTTGGATCAACCGTATTAAGCAATGCTGAGCATGGAGCCA
ATTCTGTTCACTAGATTTGCTCAGAAAGGGCCAGACGAGAAGGATTTTTC
TAAAGGCACCTACTACCAAAAAGCTGCCAAGGCGTCCAATGGAGCCCAGA
GAGAATATGCTAACAATAAAAAGTTGAACACCCTCAATAAAAAAGGGTAA
AAGTAATTAATAGAAAATTACTGAAAGCTTTTTTGAAACCAAAAGTAGTC
AGCATTGGTAAAAGTCTACAAAAGTGGACACTTTCATATAATGTTGGCAG
GAGGGTAAAAAGACATAACCTTTTTGGAGGACAATTTGGCAACAGAGTAC
CAAAAACCTTACAATTGAAGAGAACTTTGGCCTGAGTGCAGTGGCTCACA
CCTGTAATGCCAACACTTTGGAAGGCCAAGGTGGGAGGATTGCTTGAGCC

FIG. 3 (50 of 52)

CAAAAGTTTGAGACCAGCCTGGGGTAACACAGTAAGACCTCGTCTCTATG
AAAAATAAGAAAAGTTAGCTGGGCATGGTGGCATGTGCCTGTGGTCCCAA
CTACTTGAGAGACTGAGGCAGGAGGATCGCTTGAGCCTCGGAGGTCAAGG
CTGCTGTGAGCCATGTTCATGCGACTGTTCTCCAGTCTGGGTGACAGAAT
GAGACCCTGTCTCACCAGAAAAACAAGGCAAGAGAGAGAGAGAGAGAGAA
GGAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGATGGAAGGAAGGAAA
GAGAAGAAAGAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA
AAGAAAGAAAGAAAGAGAAAGAAAGAGAAAGAAAGGGAGAGAAAGAAGGA
AGGAAGGAAAGAAAGAAAGCAAGCAAGCAGGAAAGGAAGGAAGGAAGGAA
GGAAGGAAGGAAGAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAA
GAAAGAAAGAAAGAAAGAGAAAGAAAGAAAAAGGGAGAGGGAAAGGGAAA
AGAAAAGGACAAAGAAAAGACCTTTGAACCCTGAATTTCACTTTTAGAGA
TTCATCTTAAGGAAATTCATTCCAATAGAAATTTATCCCCAGGATTATCT
AAATATTTGCTTTTATTTTCTTCTAGTAATTTTATGGTTTAACTTTCTCA
TGTTTAAGCCTTTAATTTATTTGGAATTTATTTTGGTATGAGAAAGTGTG
ACCTTTTTTTGTTTTACTTTAAAAAAAATGTATTACGATTATTATTTTAG
AGACAGGGTCTTGCTCTGTCACCCAGGCTAGAGTGCAGTGGTGTGATCAT
AGCTCACTGCAGCCTTGAACTCCTGGCCTCAAGCAATTCTCCCTCTTCAA
CTTAGGAGTAGCTGGGACCACAGGCATGTACCACCATGCCCAACTAATTT
TTTTTATTTTTTGTAGAGACAGAGTCTTGCTTGTTGCCCAGTCTTGCAAT
GTTGTCTCAAACTCCTGGGCTCAAGTGATCCTGTCGCCCCAGCCTCCCAA
AGCACTGGGATTACACGTGTGAGCCACTGCGCCCAGCTGCCTTTTTATTT
TTTAATTTTTCAGATGCTTTGTTGGTTCCAAAATAGCACTTATTAACCCA
CGCTTTCCCCCTCTGGTTTTAAATACTGCAAGTTTGGCTTTGAAATACAA
CCCACTGCCTTATTCAGGCTACATTCAAGGAAATCTGAGACCAAGAGTCT
GAAGGCCCAGTTTCCTTCCTCAAACCCAGGAGGTGGTAAATGTGTCACTT
CCACACTTTCTATCTATTTCTAAGAACTCCTTCTTTCCAAACTCTGACAT
GCCCCTGGCTCAGGTCTATAGAAATTCCCAGGGTCCACAGACAAAGCAGA
ACTCACTTATGGGGAAATCTGGGAAATACTTATCTGTTAAACCTGCCCCA
TATGGTGACTCAGATTGTCTAAAGCCCAAAGCATCATTTTCCACCCCAAA
CCATTTCCTCCTCCAGACTTCTCTATTTCTGTGGTCCAGAGTCAAGATCT
TGATATTACCCTAGAGTCCCCCTTCTGCTCTCCTGCATACCCAGATGCCC
CTCCCTCCCCAGATCCATTCTCCCACCCTCCCTCCCATCAGTTTGGTGGG
CCCATCACCGCTTCCCCTGGCCCAGGCTCTCCTTTTGTGCGCTTGGAGCA
GCAGACTGATCTCCCAGCCTTCACTCACTTCATGTGGTAATCTGTTGTGT
TCATCACTGTCAGAATCTTCTGCATCCCCTCACTACTCTGCTGAAAACAC
TCTAGTGGTTCCTCATTGCTCATTAATGAAAGTCTAGATATTAAACGTAG
AAGGCCCAGCACAATTTGCCCCTATGCCACCTACCTCTCTAATCTTTTCT
CCTTACTCTGACAGACTCTCCGTCTGTCATTTATGTATTCTTTTATTGCT
CTCTTCTACTTTTAGTATGAACTGGATTTATGGATTTTTTTAACATTGCT
TTCAAGTATGGAATAAAGAATTTTATTTATTTATTTATTTATTTATTTGA
GACTGGGTCTCACTCTGTTGCCCAGGCCAGAATGCAATGGTGCAGTCATA
TCTCACTGTAACCTCGAATTCCTAGGCTCAAGCCATCCTCCTGCCTCAGC
CTCCTAAGTAGCTATGACTACGGGTGTGCATCACCACATCTGGCTAATGG
AATAAAATATTACAATGCCTAATCTTAATTTTCAAAATTTTAAATTACAT
TGTACCTAATGCCCATGCATTTACTTTTTTCAGTGGGTCAATAGCCCTCA
CTTTGGCAAAGGTCCCAGGCCCAAGGTAAGGCCTTACTTTTTCCAAACTC
ATCTTTTGAAAGACATAAGTGCCTGTAAGTTGTACCACATTAGGTTCTAG
GAATTTTTCATCAAAGACTTTATCAGACTATTTTCCTCTAAGTTGAGAAA
GAGCTGGGGGCAGAATATGGCACTGAATGACTGAAGAGAAGGCACTGAAA
TCAGGCCAGAGGTTGCTGGAAAGAGCAATGAGGAACACCAGCAGCAATGA
GGAGCCGGTGATGATTTTGGCTTCACAGGGAGGTGTGTACCACACCGATT
TTATCTCTACGTGGATGAACCACAGCTGTCGGCTCCCTTGTCTCCAGGAC
ATCACACTCTCCACATTCCCTCCCATCTTCCGGCTTCTGCTTCCCGGGGC
CCTCATCTGCCCCATCCTGGGTGAACACTGGTCGGTCAACTGCTGGGCGT
ACCTTCCCGCTCTGCACACCCTCCCTGGCCACCCCACCCACTCTCACGGC
TCGCACTGCAGAGGAGCCGCATCTCTAGCTCCAGCCCATCTGCCTCTTCT
GAGCTCTAACTTCATGTAGGCGACTCCTGCCGGTGTTGCCTCACAGGCCC
ATCATACTTCAAAGCATTTTCCCCTCAGAACACCATGTCCTGGCTGCTCC
CTCCAGAAGATACATCTCTCAAGCACATCCCCGCGGCTCTCACCTGGATG

FIG. 3 (51 of 52)

ACTGCATTCACCTTCTC ACATTTGCCCTCCTTTGGATGTATATAGA.
GTTTTAAAATACAAATCTGATGTGCTTGCTCTCCTGCTTGAAACACCTCA
AAACTGCCTTCAGGATAAACCACTGCCCTTGACATGTTCACAGGTTGCCC
ATGGCCTGGCCCTGCCCATCTCTTCAGCCTCATCTCATGCCCCTTGCCCC
TCGCTCTCTGGGCTTCTGCCTCCCTAGCCCTCCTTTAGGTTCTCTAACAC
ACCATAGTCCTTCTAGTGTTGGGGCCTCTGCAAGTGCTGTTCCCATTGCC
TGAGACATGAATCCCTCTCCCTATCTCTACCTGCACCTTCATCTGATTAA
TGCCTACCCTTCCTACTCATGATGTTGCTTTCTCAGGGACTCTCTCTGAC
TTTTTAAACTAATCAGGGTCTCCCCAGTATATATCTTCATAGCACTCTGT
ATTACTCCTTTCTTAATGACCACCTGCTGTAGACTGAATGTTTGTCTTCC
TCCAAAATTCATATGTTAAAACCTAGCCCCAAATGTGATAATATTTGGAG
GAAGGCTCTTTGGGAGGCAGAGCCCTCATGAATGGGATTAGTAGCCTTAT
AAAAGAGACCCCTGAGGGCTCCCTTGTCCCCTCCACCGTGTAAGGATGCA
ACAAGAAAGTATGGTCTATGATCCAAAAAGCAGACCCTTGCCAGGTACCC
AATATGCTGGCACTTGAACTTCCCAGCCTCCAGAACTGTGAGAAATAAAT
TTCTATTTTTCATAAGCCACCGAGTCTATGGTATTTTGTTATAGGAGCAC
AAACAGACTGATGTGCCACCCAACCATGATTATACGTGTAATTTATGGTT
TCTCTGCTAGTAGGGATGCACCATGGGGTTAGGAACCACGCTTTTCTTAT
TTCCCACACAGTCCTTAGCTCTAAGCATGTTCCTGAATCAAAGATCCCCA
TCTTTTATGAATGAAGGAGTCAGTGAATGAATTAATGAAAGAACTGATAA
CCCTCAATAATTATTCCAGCCTTTTATACCTACTATTAACAAGCTTGCAT
TCTACTCCAAATTTATTGGGCTTTAACTCTATTTTTGGCCAGCCACATTT
GACATTCCCTGAAGTAAATCTATGCTTTCCATCCTAAGTCAAGGAAGGAC
CTGGACTAGTAGGGCCAAGAAAGGTCTAAATTCCATGGGTGGGAGAGAGA
GACTAAATCTGAAAGGAAGAATAGATTGAGCAAAGGTGTAGAGATTGGGG
AAGGCTGGACATTTGGAGAGAAGGAAAAGGAAACTGACACTAAACCAAAC
AGTCTCACAAACACAATCTCATCCTTCCAAAACTCTGTGAAGTAAGAATT
ACTATCCCAGGGCCAGGCACAGTGGCCCATGCCTGTAATCCCAGCACTTT
GGGAGGCCAAGGTGGGTGGATCACCTGAAGTCAGGAGTTCAAGACCAACC
TGATCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTAGCTGG
GCATGGTGGTGCACACCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGG
AGAATCATTTGAACCTGGGAGGTGGAGGTTGCAGTGAGCAGAGATCGTGC
CACTGCACTCCAGCCTGGGTGACAGGGAGACTCCGTCTCAAAAAAAAAAA
AACAAAAAAAAAAACCAAAAAAAAAAACAAAAAACAAGAATTACTATCCCAG
TTTTGCAGATGAGGCAATGGAAGCTCTAAAAAGTTAAGTAGGAGAAACAA
ACATGAAATGTATGTCTTATGCTTTTCCTCATCCTATTTCCTCAGCCTGG
AATGTCCATTCTCCCTCCACTATGCAAATCTAACTCTTCAAGCTAACACA
TAGCAATGTCTGAGAAACCGTCCCTGTGTTCACTCTGTTAGCCTCACTTG
CTCCCTCCCCATCCCTCTGTTTCCTTTCTGTTATAACACTTCTCTATTCT
GCTGGCATCACAGTCATCTCCACCTGCCTTCCTCACAAGTTAAAAGCTTG
TTAAGGGCAAGTGGTGTTCTTTGCCACCTCATTCCCCAGGGCTTCTAACA
CAGTGCCTCATGCATGACAGAGTTGTAAAACAGGTTACCAAGCTGGCTTC
AGGCAGGTTTGCATGGAACTGTGCTTTACAGGAATACCTGCTCCCCCCAG
GCCCTGGGTCTTCCTCCTGAGTCCAGGCTCAGACTCTCTCATCCTGCTCG
TTCTCTCTTGGGGAGCCACAGTAACTTTGAGCAACTTTGCATGGGATAGA
ATGGCCTATTAGGGGCAGCACAAAGACCCCATGGAGGGAAGAGTACAGAA
AGGGAAAACGATAATCATATTTTTTTAAGATGTGCATTTTCTTAACAAAA
TGCTCTAGTACTTGTCCAGACTTTCAAACTCAAAAACCTAAGCGTCCTTT
TCTTGAAGATCATCAAAGGCCCCAGTGGTCCTTCAGGTATGTCAAGCTTT
CTAGAAAATAAAGGTAAGTCATAATCACTTAACACACATGGCTAAATGGC
CATTTCCTTCTAATTTATCAGCAACTGTTACATATTTCTATACTAGAAAA
AATTTATATTTATACTCAGGGTGGTAAGTTAAATTTGCCATCGAAGTAAA
GCAGAAAGAGCGTAGCATGTATGTATATGTAACTCAACTGTGCATGAGAC
AAAGATGTCTTGAGGAGAATGAGTCTAAGATGCGCCTGAGCAATAGTACC
C

FIG. 3 (52 of 52)

>Contig1
GCACCCATGTTTCTAAAGGGCATACCAGCCATAATAACAGGATGGGTGAG
GATATAGACAGCAGATGACAGAGAGGAGAGTGAAAGCTGGGAATCCCAGC
TAAAGGCATCAGGTTTATGGAATGAGTAGGGGACAATACTGTGTGTGTTT
ATACACACATGTATATGTGTGTATATGTATACATGTTTATGTATATATAT
AATTATATGGTACCATTTCTAATTGACAAAATAATCTATCACATTTTACA
TTATCAGATTTTACATCTATTGTTCTAAATACACTCAGTCATCAGCCCTG
TGTGTGGGCTCTTACCCATCCCCATGCACACCTCAGCTCAACCACTGATG
GATGGATCATCTGCCTATCAGAGGTGGCATATTCAGGTGAATCCATGGCC
ACAGCTGCAGCACTTCCTACCCACGCAGAAAGGCTCCACAAGAGGAGGCA
CACCCGCTCTGACTGTCCCTAAGCTCCTGACATCTTCACCCCATGAAACT
GCTGCTCCTGGGTGCTTCCTGCCTTGCCCTGCCCACCCTTGTACTGTTCT
CACCATTGACACAGCTGGTGCCCGATGCAC
>Contig2
NAAAACGAATCGTCACTATTGAAGCCTGTCTCTCANCGGATCGTGACTAA
GAACCCCCTCCTTGCTTCAAGTTGTCCTGCCTTTCTAGGCAGAGCCACCC
TACATCTTAAATATATTGATTGATGACTTACGTCTCCCTAAAATATATAA
AACCAAGCTGTGCTCTTACCAACTTGGGCACATGTGGTCAAGACCTCCTG
ATGCTCTTGTCATGAGTGGGTGGGTGTTCTCAACCTTGGAAAAATAAACT
TTCTAAATTAACTGAGACCTGGGTCAGATTTTTGGGGTTCACAGCAACAA
TTTAAAAAACTCACCATTGACCTGAAATTTTGACCTTATGCTGTTGCTCA
CACTCCTCCATGAAAATAGACGCCATCCTATGAGTTCCCTCAGCCATGTC
ATGCCACACTTCCAACATGTGTCCCCATCCACCATCTGTCTTCTTATTGC
TGCATCCTACCCAGGCCCTGATCTCTGGACCCATTGTTGTATAATTAAGA
ATTTGGGGCTGGGCATCGTGGCTGTGGCTCACTCCTGTGATCTCAACATT
TTGGGAAGGTGTATTAGTCAGGATTCCTCCGAAGGATGCAACCCTAGGGA
TCCTCTCTATGACCCTATGTCTA
>Contig3
CGCGCTCAACCGACCGATTTGCGCGAACCTGCCCATGCCCGAGGACAGTG
TAATCCTAAAACGTCCCCTGAATCATAAGGATATGAGTGCGAAAGTACGG
TTCCCTCTGTCACCACTTTCTAACAACGCTATGTCCGATCCGTGCACTAA
CCCCGCCCAAGTCACTGAAACACTGATGGGCGCTTCCTCTACAGGTATCC
AGGGCCAATACCACTACTCCCCTCCTCCCTGTCCCCCTTCCACTCTCTAG
AGGCCGCGGATGCCATCCTCTATTAGCACAACCGAAAACGACGGTGAAAG
TACCACGAAGCTCACGATCTGATCGGTCGCCCAATGCGGTTACAACGGCT
GTCATCCCAACCCCCGTCCCATCCTCCATATTGCCCCCCCCTATGAGGAT
GGCCCTATCATCATGACCTCCAAAATTCTGTCATCTCCCGACGTAATGCC
GCCCCTCGAACGCCTGACACCATCAAGTCNGTCACCTCCCAAAATACTCC
TCCTAATCACCAGGCCGAGTATCCCCGGTTCCACAATACCTCCTTGAGAC
GGGCCGATATCACACAC
>Contig4
NGGAGTTTAGGTCAACTAGTAACAAGTGGGATTTGCGACTCAGGTCTATC
TAATCCTCAAACCCACGTCCTGGACCCCTACACAGACTGCCCTCCCTCAG
TCCTCTGTGTGGCCTCAAGAAGGGTCTGGACATTCAAGTTTAAAAATCCA
TCCAAAGAATCTATGGACCCAGTGGTCTCTGGAGTCAATGTTCTGAGGCT
CAGAAGGGCCAGGCAGGAGGGAGCCGCCTCTACACAGTCCTGAGCAGAGT
GGGCTGTGTCCCGGCACAGCAGGGGAGATCATAAACAGAATTCTGCCCTG
GGCCCTATTTAAGTAGGACCTTTAGGCTGCCGGTGTCATGACCACAGGTC
CCANGTCTGCACGATTGGCTGTGTGTGGAAAATCTTCACTCCTTGCGGCC
TTGTCCTTGGCAGAGAGCACCGCTGCTTCCTCTGATGGCCACCAGGGGGA
GGCGCTCCCCTGGGAACGGTTTGAANGGGAGCCTCACCCCACACGTGCCT
TCCGTGGTACCCAGCACCAGCTGCTACCCATGGTTACCCACAGGCCCAGC
TCTGCTCTGAAGAAGGAGGAGTGGTGGCGATCANGCCTTGTCTGCATCCC
GTGGCTGCCCCTTTCTTTTCTTT
>Contig5
GGGAGCTAACCGCTCACTGGGATTACAGGTACGCACCACCACGCCTGGCT
AATTTTGTATTTTTAGTAGAGACGGGGTTTCTCCGTGTTGGTAAGGCTGG
TCTCGAACTCCCAACCTCAGTTGATCTGCCCGCCTCAGCCTCCCAAAGTG
CTGGGATAACAGGTGTGAGCTACCATGCCTGGGCTTATATGTTTCTAGTC
CAAACATTTAGCTACCTTTTTTTTTTTTTTTGAGACGAAGTCTCACTCTGT

FIG. 4 (1 of 61)

TGCCCAAGCTGGAGCACAGTGGCACAATCGTGGCTCGCTGCAGCCTCAAC
CTCCTCAGGCTCAGGTGATTCTCCCACCTCGGCCTCCCTAGTAGCTGGGA
CTACAGGTACGCACCACTACACCCTGCTAATTTTTTTGTTTTTGTATTTT
TTGTACAGATGGGGTTTCTTCATGTTACCCANGCTGGTCTTGAACTCCTG
GGCTCAAGCAATCTGCCTACTTCAGCCTCCCAAAGTGCTAGGATTACAAG
CATAAGCCACCATACCCGGCCTACCTACTTTTAACTTGTGGAATTTTCTA
TAAGGTCANGGATGCCTGNGGGAACAAAAGTTTCTCCCTTGGTATATGCA
AGTAAAATCCACATGCTGCCTCCC
>Contig6
AGGACTGTAGCTGTTGTCTAGTCACCAGGCTGGACTGCTTGGCATGATCT
CAGCTCACTACAACCTCCACCTCCTGGGTTCAAGGGATTCTCCTGCTTCA
GCCTTCCAAGTAGCTGGGATTACAGGCATGCACTACCATGCCCGGCTAAT
TTTGTATTCTTAGTAGAGACGGGGTTTCGCCATGTTGGCCAGGCTGCTCT
CAAACTCCTGCCCTCAAGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTG
GGATTACAGGCGTGAGCCCCCGGCCCACATGTAAAAGTTTATATCTCTGT
TGTTTCACCTTGTTTTTGACCTAGTCTTTCAGTGATTTGAATCTTGATTC
AGTCTTTTGTTATTTTAGTGGTACTTCCCAGCTTTGTGTCATCTGTGGAT
GACATATGAGTCTTGCTTCTTCATGCCAATTTAAGAAGACTGAACGGGAA
TAGGTCAAAGGCATGGCCATGAGCGATTTCTCTCCAGCTTTTCATGGTGT
TCAGCTTCAAATCTATTCACATATTGGACCTGCAAGCCATCATCTTATCC
ACAGGCTATCATCATAGGTGAATGTAAATTGGGTTTAGGTGGCCAAGCTG
AACGTGAGATATNTTC
>Contig7
AGCATGTTCTCTAAAGGCCTATCAAAGCTGACATCAAAGGGATAAGTTCC
AGTTACCCAGCTGAAGGGAAGGAGGGTGTTTCAGATAGAGGAAGGATAAG
CATGACCTATTCAAGGCCAGTGAAAGAAGCGTGCAACGGCCAAGTCAGGA
GAACCTGAAATTGTGTCAAAGAGCTTGGATGCAAAGAGCCGTGGGAGACT
ATTGGGGGTTTTAAGCAGGGATATAATATTCATTCAAGCATGCAGTAAAA
GGTCACTGGCACCTGCCATGGGCCAGGACTCGGGCTCTACATGATTGCGT
CTGTTTTGGAAATATCACCCTGGCTGTGAGATGAAGAACAGGTAGGAGGG
TCACAAAACTTGAAGCAGAGAGACTGTTGAGGAAGTAAGCTGTTTTTGTG
TGGACTGTGGCAATCACAGAGGCAGAGGATATAAATGCACAGAGACACAA
GGCATGTGGGAGGCAGAAGGAATCAAATACAATGAGTGATCAGATGTGGG
GTTAGAATGGTGAGTGANAAAGACATACTCAAGGTGACACGCCCAGGTAT
CTGGGTGGATGGTAAGACATTCATGGACTAGAATCGAAGAGGAGGTGGGG
ATGGACATTCCTTCCGTTTAGAGGGGTTCACCAGGAGGATTTGCCGGAAC
ATGGAGAGGATTAACCAGGAATCCGGTGCCTTTTTCCAAACTGGGTTGGA
GGGG
>Contig8
GGTGAATGCTTTGGCACGCTGTGTAGATTTTAGGTGACGGGTGGTGACAA
TGAGTCCGTGTCGAGCGCTGATTTTTTCGGCCTTTAGAGCGAGATTTATA
CAATAGAATTTGGCATGAGATTGGATTGCTTTTAGTCAGCCTCTTATAGC
CTAAAGTCTTTGAGTGACTAGATGACATATCATGTAAGTTGCTGATAGGT
TTCCAGTTTTCCGCTCCTAGGTCTGCATATTGTACTTTTCCTCTTACTCG
ACTTAACCAGTACCAACCCAGCTTCTCAACGGATTTATACCATGGCACTT
TAAAGCCAGCATCACTGACAATGAGCGGTGTGGTGTTACTCGGTAGAATG
CTCGCAAGGTCGGCTAAAATTGGTCATGAGCTTTCTTTGAACATTGCTCT
GAAAACGGGAACGCTTTCTCATAAAGAGTAACAGAACGACCGTGTAGTGC
GAATGAAGCTCGCCATACCATAAGTCGTTTTTGCTCCCGAATATCAGACC
AGTCAACAAGTGTCAATGGGCTCGTATTGCCCGAACAGATTAAGCTAGCA
TGCCAACGGGATAAACGAGTCGCTCTTGGTGGAGGG
>Contig9
GGGGTGGGGCGCCTGGTGTTTCTAAAGAGGATCTCCTGCCAGAAATGGTG
TGCTGACACTGTTGTCCTCCTTGGTGTGGAACTTTGGTGGGAAGAAAGGT
TGGAAAGGGAAATTTTGATCCTTGGATTTAACCCGAGTTTGTTACTGATG
CTCACAAGACTAGGGGAAGGATAAAGGCAGGTGAGTCACTCTAGGATGGC
TCANTGAGCTCCACAGAGCTGGAACCACAGGCACCAGGAGGGATTCAGAG
CAGGCCTCAGTGCACGTCAGCTGAGTGAACCAATGAGCAGGTGATGGGTC
CAGGCAGAGCCCTGTCCTCTTTAGGCAAAAACCCTTGAAACACCGTTCCC
ATCCTAGCCTGTGTTCCACCCAAAGCTGGCCAGTCTCCAGGCCCTGCCTG

FIG. 4 (2 of 61)

AGCCCCAAGGAAGTGGTATGGTGAAACAGAAGGGCCATTCCTGTCCAATG
TGTGAGGAACTTCATTTCAGACTTGTTGGAAGCCCTGATGTTCAAAAACC
TCAATGATATCATTCATTTTCCCCATCCATTCAATGCCCATCCAATGCCC
ATCCGTTCAATGCCCCTTCCATTCCTCTTCAGGGAAATGAAAATTGTTCA
GAAATCCTTTCTCTTTCGAGAAACCAACCAAACCAAAACCGCGAAATTCA
CTAAACTAGCCAAGACACAATCCTGGGTTATTTTCCTTTTCCCAAACCTC
CTGTGTTTAAATTAATTCTACCCTGGTTCTCGGCCCTTACTGCGAAGGTG
AACTCACCTAACCTCTCCCAAACAGAGAAGAAACTTCTCTTGGTAAAATG
GGTTTTAACACTTCTAAAAAACCCCC
>Contig10
GCTATGGTTCTAAAGGTAATGGACTATGGCGTACACAACGTCTCGCTCAT
CGTCTGCCAGGAGGCTAAGGTATCCACGGACAATCGCTGAGCAACAGTGT
CGTTGATCCATCTCTGTACGCACTTGTCAACATGGCAGGAGTACGGGAGC
TGCGAGAATCCTCTCTGCTGATGTCCCACGGAGCATGCCGTGAGACAACG
CCACGAACGGCCCTCGGAGANANCTACTCTGCAATGAAGACGTACGATAC
ACACGTAGGAGTCCTAGCTCACCAGCCGTATCTAGGTATACTGTACTCGC
GGATACTCACTCGTGCATGCGGCAATAGATCGATACGCAGTCGTCACGCC
CATGCTCTCAGTGTGTGACCTTCTGGCGGTAGCGTNGTGGGCGCTATTAC
TGTGCGCAGCAGGCGCNTCGTACATGTGTCGGGTAGCGATGCCAGGAGCT
GTAACATAGCAAGTCGCCCCCCTACTCCTATCACTATCCCTACGCTGGAC
CGCACTCGAGATCTGAACGCACGTCTTAACCTGCCAGTACTCGTGAGACC
TATACTGCGCAAGCCTTGGCTAGGAGATCCTGCAGCGCCGGCAAAGAATC
AGCTATGATCCCCTTGCGATTATCGCACACGCACCATAGAGTATGTGCAT
ATTAACCTCTGAATGTGCTGCAAGCAGACGGTTGCTCAACATATATATGG
ATGTGGGGAAATCGCCCTGGTCACCGCCACTTGGCGTCAGGAGGCACCAG
CACGTCTGAGTGTCACGCACGTTACTC
>Contig11
GGCCGAATGGTGAATTCATCCGTCGTCTCGAGGGGGTGAAAGACGGGGAG
TTATGCTGTAATGGCACCGCTCACCCTGGGCTTATGAGCAGACCTAACCC
TCCCANAGTGCTGGGATTACAGGCATGAGCCACCGTGCCCGGCCCAGTAT
CTGAACTTCTGTGGCCAGGCAGAAAAGGTCCTGTGTTACTCGTCTCCTTT
ATCATTCATGTCCATATTCTCCCATTTGCTAACATTTATGTTTCTGCTCC
ACTGGATTCTTTGGATTTTTCTAGAACATACCCATGCTTTGCATTGCCTT
GGTCTTTGAATATTTGGTCCACTTTTCCTGCAAAGTCCCCTCTCACCTTA
TCTTCCTGGTAAACTTCCAGCCAACACCTCTTTACTAACCAGAGAAACAT
GGTTCAACTGTGCACAGGCTTGCACAGAAACTGTTCTCATATTGTCTTGT
CATTGTCAATGTGGCAGAGATGCACCTTAGATACCTCTTTGAGAAAGGAC
TCACTGCCCAGCTGCCTGGCACGTGATGAGCTGATAGCTCCAGCTATAGA
CTCCTTTAGGGTCAACCTCTGCTTTCCAGTTGAGATCATATCCTTTGCAG
GGTGGCCTCCCCAGTGATGACTAAGGCAGTGTTACAATGGCCTAGTCATT
TCCTCCCAATGCTGGACTCCCAATGAACCATCTGCTCCGGAGCTTCCCAC
TGGGCAGTCAGAGACCTTAGCTAGTCTGCCTCCGAATCAGAAGGCTCTCT
CTTGCCACTCTGGCC
>Contig12
GCTGTGTCTAAAGATTCACGGCTGTAGTTCCAACTCCCGCCGCCCTCTAC
TGTGTCCTCTTAATGGCAGTCATTCACCATCTTCCTGTCCCTCCCCTTCA
TTTCTTGGATGGTGACTGTCACTTTGCTGCAACAGAACCCTGTCCCAATC
CTTGATGGTTCAATACACACATAGACATTCTTTTTAACAGGGCGGCCTCT
CAGGTCTTTAATTTTCTTCCCTCCAATAACCTTGTGATGATCCCCCAGCT
TAGCCACTTACTGCCAGATCATTACCAGTAACTCCAGCCCCTCCTTAATT
CTAGTTTCTAATATCCTAATCTGTGACCTCACATTCCAACTTCTTCATTC
TTATCCCCTGAGTCAAAAAAATCCTTTGATCCATGCAATCCATTAAGTCAT
CTACCTTTTCACCATTCTTCGCCCCACTAGGGTTCTCATTCCTTTATTAC
CCATATGAAATTCCAAGGCCTGTTGGAATCACTCCCTTGCAGCCACTGTC
AATACTTCTGCCCCTTTTACTTCATCACCCTTATGTGGCAAAACCACAGC
CCTGGTGGAGTCGATCCTTACCCCTGCTCTGTGCCAACAGCCGCACACGC
ATGGCTGATGGAGGTTGGAAAAATCCACACATGCAGTGGGCCCTGTATGT
CCATATACGTATCCAACCTCCAGCCTTGCATATGCCTCAGTGCTGCCTGA
CAACACATTATATGTTTTCCTTAGTTCCTTCAGTCTCCTGGGTGCCTAGG
TGAGTATCTCAGACATCCTTCTCTCTCTGCAAAGCTCCAACACCTCCACG

FIG. 4 (3 of 61)

TCACATTCAACTGATGACTGTGTCTCCTATGTCACTTAGATCACAGAGGC
ATACATAAACAAATCCCAGCCACTGCCAGCACTCTGCACATCTGCGAGCA
TGGCACCCCCAATCTAGGCCTTTCCTGCTGTCACTTGGGGTGAGCTGATT
ATACTCGATCCTAGTCATTTCTACTTATGCAC
>Contig13
CTTAAGGCCTCCCTCTAACATTTTAATTTAAGATTGAAAAAGCAAAGATT
ATTCTGTTTTGGCTGCGCCTATAGTAAAGTAACCCCTATGNCAAATTTTG
ACACCTTATAGTATTTGACAGGGATAAGTATAAAATTGCTTGATTGATAC
ATCCACACCCAAATGTATGCTGGGAATGATTTTGTTTCACGGCACTCATT
ACTTAATTTTTAAAACTCTTATTTAAATTTGCAATGTTTTAAATGACCAT
CACTTAAAGTAGTAATCAACAGAGGTTAGGAGAACATAACAATACTCTTT
CTCTTAGAAAATACAACAGAAATATAATTTTTTACAGTTTTGCTCCCAAA
CTTTTCTCTGTAATAACATGCCTTACTCACCTTTACAATAGGTTTGTTGT
GAGAATCTTGTAATGTAAACCCTGGGTGTTCTGTGAAGCATTTTTAAACT
TCTAGTTTACACTGACTCTTATTCAAGTGTTTTTAAAAATATATTTAAAA
AACTGGCCAGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG
CCAAGGCGGGCAGATCACAAGGTCAGGAGTTTGAGACCAGCCTAGCCAAC
ATAGTAAAACCTCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGT
GGCGGGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGAAGAATCG
CTTGAACCCGGGAGGCAGAGGTTGTGGTGAACCAAGTTTGCGCCAATGCA
CTGCCAGCCTCTGCAGNGACAGCC
>Contig14
GGGGGCGGGCCGAGTGATCCTAAAGCCCGCTCGCTTCACAACAAAGCCTA
ACAGTCCAATCACTTAATGCTGCATTTATTCCTGGGGAAGCAAGTCTCCT
TTGCACTTTACACAGTGAGATAATCAGTTTCTCATGTGGACCACTGGGCC
AGGAGGGCCTGACAAAGGGCAGTCTACATTTCAGACTGGAAACTGCTCCC
AGAACTATTTCTTTCTAGTTCCCACCTCGGTCTGAGGTGCCTGAGGAGAG
GGACTCAACAGAGGAAGCAGGAGCATAGCTCAAAGTCTCAGAACATGGAA
GAGGAAAAGAATCCTCACAAGATTACGTAACTTACAGGCGTGTTGCTGCT
TCAGTAGAAGTTTCATCTCCCTCAATCCTGTACACTTTTCCATACATTAC
ATACTCAAACTGGTCAGCCCTATGGAGCAATAGCAGCAAAGTTATTCTTA
ACAGTAATTAACAATATAAAAGATCCCATTTAAAAATGGTTACTGGTCAG
CCGGGCGTGGTNNNTCNANCCTNTAACCCCANCACTTTGGAAAGCATGCG
GGCGATCCCAAGTCTGATATCGAAACATCTGCCTAACATGTGCAACCCCT
CTCTACAAAATACAAAAAATATCCGGGCTTGTGTTGGCGCCGTTATCTCA
CTACCCGGAGCTAAGTAAGAAATGCTTTACCTGGAAGCGATTTTTTTACT
TATATCCCCTCTCTTCACCGGGCGCGACCAAATTCTTTAGTATAGGAAAG
TTTATTGTTTTATGCCTTTGTCAAGGCTCTACTGTATCTTTTCTGTCCAC
TCAC
>Contig15
GGTTCTGAACAACAGCAGGCGATTCCTAGCCCTGTACCCGGGGCATTGTC
CAACACTCGACAGGGCTGAATTCGTCCATAACGGTGTGCCCCTCTGGGAT
ATAGGATGAAATGAATTGATCTGAGTACCTGGGATGTAAAGTTACTAAAA
CGCCAGCTAGGTTCACGCCCCGATGCTTAAATATGATCGTGGCCTACACC
TCGTCCAGCAGAAAAAGTACCCTTTCTTCAACACCACCTCACGATCCTCC
AATTTAGGAGCTATAAAACTCATGACTCTTTATTTACCCCCCTGCAGATTC
TCAATCCAATAGTGTGTGTCTCCCTGTGAACTCACGGATATACCGATTTT
CCCCACGTCATTTCCACACGTCGCAATCGCTTAGTCATCCCTATGTATGA
GAATCATGGATGACTATGTTGAAGTCCATCTATAAAGTTCAACCCCCATC
TCCGTCCCTGATTCCCCCTCCCCAAGATCACCAACGCGACTCGACATATT
GTTATCGCCCAAGGGACCTCTTGCATCCCCCATATCCACTGGTCACCTCC
CCTCTTGGCTGGAAGTCACCGGGAAGTTCTCCACATGTTGT
>Contig16
TGCGAGCGATGTTCCTAAACTTTAGCGCCATTGACTCGAGCATGGTCATG
GCTGTTTCCTG
>Contig17
AGGGTGTTCCTAAAGGATACTACGTTCCCTAAAGTCCAGAGAAAAAAAAA
AAAGTAACATAATGTGGCTTATTTGGTATAAAAATTTTACAGGAAGCATT
GTCAAATATGAAATAGTGTTTGGTTTTGTTTGGGCTGTATTTGTATAAAT
ATGTTATTGGTATGTGTTCCAAAATTATAGGAAACTCCTATAATTCTGAT

FIG. 4 (4 of 61)

ATGACTTGGTGTACATTATCAGTAATAATTATAATTGTTATGGTAAATTA
TTGTGTGCCATGGAGGTAACAAATTTCCTCATCAAGTGTGTCTTTGACTA
TGGTTGCCCTAAAACTTTTTGCCATTCACAGACAATTGTCTTGCTTTGGT
CCTCTTTAGAAGGTGGTTTTATAATCAGCTATAAAACTCTAACGGGTGCT
CTTGAATGCAGGCTTAAGATAGCTTTGGAGACTGTGACATCAGAATAGAG
GAAAAACTTTCAGTATTCATGGAGTGCTGAAATATTCATGAATATCAAGC
AAAACAGGAATTAACTTCATAGATGGAACTAAAAGAATGCTGAAGTAATC
TTTTTGACTTTTTTTCTTAAAATGTTGATCCTTCGTTTTGTTTTTCAGAG
TCAAGGAAATTTTTCTGTTGAGATATTGACAGCTTTTAACAATTAAGTAT
ACTCCAGTGAACACAATTTGGAGCATATTTGTGTCTCTCTATATATATTT
GGAAACAATNTTTGAGTATTCTTAACTTATTGCAATATT
>Contig18
GGTTGTCTGCTATACCAGTAATGGGATTGCTGGGTCAAATGGTATTTCTG
GTTCCAGATCCTTGAGGAATTGCCACACTGTCTTCCACAATGGTTGAACT
AACTGACACTCCCACCAACAGTGTAAAAGCATTCCTATTTCTCCACATCC
TCTCCAGCATCTGTTGTTTCCTGACTTTTTAATAATCGCCATTCTAACTG
GCATGAGATGGTATCTCATTGTGGTTTCAATTTGCATTTCTCTAATGACC
AGTGATGATGAGCTTTTTTTCATGTTTGTTGGCCACATAAATGTCTTCTT
CTGAGATGTGTCTGTTCATATCTTTTGCCCACTTTTTGATGGGTTTTTTT
TTCTTGCAAATTTGTTTAAATTCCTTGTAGATTCTGGATATTAGCCCTTT
GTCAGATGGATAGATTGAAAAAATTTTCTCCTATTCTGTAGGTTGCCTGT
TCACTCTGACAATAGTTTCTTTTGCTGTGCAGAAGCTTTTCAGTTTAATT
AGATCCCATTTGTCAATTGGCTTTTGTTGCAATTGCTTTTGGTGTTCTAA
TCATGAAGTCTTTGCTCATGCCTATGTCCTGAATGGTATTGCCTAGGTTT
TCTTCTATGGTTTTTATGGTTTTAGGTCTTATGTTTAAATCCTTCTTTTT
TTTTTTTTTTTTTTTTTGAGATGGAGTCTTAGTCTGTTGCCCAGGCTGGA
GAGCGAGTGGCGTGTCTNTAGGACGC
>Contig19
GCATGTTGTCTAAAGGTTTGTCTTCCTCCAAAATTCATATGTTAAAACCT
AGCCCCAAATGTGATAATATTTGGAGGAAGGCTCTTTGGGAGGCAGAGCC
CTCATGAATGGGATTAGTAGCCTTATAAAAGAGACCCCTGAGGGCTCCCT
TGTCCCCTCCACCGTGTAAGGATGCAACAAGAAAGTATGGTCTATGATCC
AAAAAGCAGACCCTTGCCAGGTACCCAATATGCTGGCACTTGAACTTCCC
AGCCTCCAGAACTGTGAGAAATAAATTTCTATTTTTCATAAGCCACCGAG
TCTATGGTATTTTGTTATAGGAGCACAAACAGACTGATGTGCCACCCAAC
CATGATTATACGTGTAATTTATGGTTTCTCTGCTAGTAGGGATGCACCAT
GGGGTTAGGAACCACGCTTTTCTTATTTCCCACACAGTCCTTAGCTCTAA
GCATGTTCCTGAATCAAAGATCCCCATCTTTTATGAATGAAGGAGTCAGT
GAATGAATTAATGAAAGAACTGATAACCCTCAATAATTATTCCAGCCTTT
TATACCTACTATTAA
>Contig20
ACGGTTCTCTAAAGACTTTCAAGAGCTGGATTTTATGCTTTAGGTGAAGG
TGATAAAGTAAAGTGCTTTCACTGTGGAGGGGGGCTAACTGATTGGAAGC
CCAGCGAAGACCCTTGGGAACAACATGATAAATGGCATCCAGGGTGTAAA
TATCTGTTAGAACAGAAGACACGAAAATATATAAACAATATTCATTTATC
CCATTCACTTGAGGAGTGTCTGGTAAGAACTGCTGAAAAAACGCCATCAC
TAACTAGAAAAATTGATACCATCTTCCATAATCCTATGGTACAAGAAGCT
ATATGAATGGGGTTCAGTTTCAAAGACATTAAGAAAATAATGGAGGAAAA
AATTCAGACATCTGGGAGCAACTGTAAATCACTTGAGGTTCTGATTGCAG
ATCCAGTGAAGGCTCAGAAAGACAGTACACAAGACGAATCAAGTCAGACT
TCATTGCAGAAAGAGATTAGTACTGAAGAGCAGCTAAGACACCTGCAAGA
GGAGAAGCTTTGCAAAATCTGTATGGATAGAAATATTGCTGTCGTTTTTA
TTCCTTGTGGACATCCAGTCACTCGTAAACAATGTGCTGAAGTGGTTGAC
AAATGTCTCAAGTGGTACGCAGTCATTACTTTCAAGCAAAAAAATTTTAT
GTCTTAATCTAACGCTATAGTAGGCATATTATGTTCGTATTATCCTGATT
GAATGTGTGATGTGAACTGACTTTAAGTAATCAGGATTGAATTCCATTAG
CATTTGGTACCAAGTAGGAAAAAAAAATGTAAAGCCAGTGCTTAGACACA
GC
>Contig21
CGCTGTCTTAAGAACTGGGCTAGGAGTGAGCAGTGAGCCAAGATCGCACC

FIG. 4 (5 of 61)

ATTGCACTTCAGCCTGGGCAACAAGAGCAAAACTCCATCTCAAAAAAATA
CATATATATATATGACCCATAAAAAGGAGATAAATCAACACTTCAGAACT
GACCCAAACTTGCAAAGATACTATAATTAACAGAAAAGGACAGTTTACTA
AGTACTCCGTATGTTCAACAAGTGAAAGATTAAACATATTAAGTAGAGAT
GTAGAAGATATAAGAAGATCCAAAATGAACTTTTAGAGTTGAAAACTACA
ATATTTAAGATAAAAATACACTAGGTGGATTAAAAGTAGATTACACATT
GCATAAGATAAAAAAAAATGAGCCTGAATACAGCACAGTATAAACTATCT
TAAACAAAAACACAGAGAGAAAAAATAACTTTAGAGACTTAGCTCTTATC
CTCTATTTGTTTCTAAACAGAGGATAAGGGGCAGAAAAAATGTTTGAAGA
AATCATGATTTTTAAATTTCCAACTGAGATAGGAATAGCACTGGGTAGTC
ACAGGAGGCTGGAAAGACCCAAACAGCAGTTAAAACAGGAACTAGGCAAA
GAAACCAAAGGATAACAGTAAACCTAAACTAAGGGAGAGAAAACTGACAA
AAGCTGACTTAGGATAACTGAC
>Contig22
CCTGAATATAAGCCGCAAGTAACCAATTAAATTTGTTTTCCAAAATTGTA
TTAACAATCTATGAAATTTTTATCTTGACCATAGCTATAACTTCCAGAAG
CCTTTTATAACCTCTATAACCTTTATTAAGGAGTAGGTTAATGCTTCAAG
AAAACCTTGTTAATCTGACACAGGACCCATATGCTGATCTTGCATCAGTG
TGGCTTGGACATCAATGATTATGATTAATTTATAGAGAAATTGAACTTAT
TTTATCTCTCAAAATTGGCCCTTACAATCTCACACACCCACCTCTTCCAC
TATAGTTCCTGGGCCTTGAGTTGAATAGCTTTAATTTCTGGCTCTGTGTT
TCAAGAATGCAGTTTATTTTGATTGGCATTTTCTACCAGTCCTGAAGATG
AACCTTTAATTGCTGTCAGTATTTAAGATTTAGCAGGACTTGTCCTTTTA
AGAACCAGGAGTCAAGCCCTATAACTCAATGTCACAAGGACTTTAAAAGC
ACATACATAAAGATATATGGATGTAATAATCATAATTTTTAAAAAATTGT
ATTAATCTCAGTGTTTTCTAAGCAAACCAAAACTTAATAATAATGGCATA
GAAATTATTTCAATAAAACATAAAATCTGTTAAGCCAGTTACCAAAAGGC
AAAAGAAAAGACCTTCTGCAATGCACAGAATATTATGTTGGAAGAAAACA
TTTCCTTTAGACCTTTAAGAAAACATTGTTAGCATCAGGACACAACAAAC
AGAATCTGAGGGTAAAAAACGTATATGAGCTGAAGGGAGTTGAAGGAGGG
CATTACTATTTCCCACCCTTTTAAAGGGGAGAGAAAACCTAAAACAGCAA
GATGCAATAAAAGCTGAACTTTGGGTTAAAAAAAAATTCTTAAGTCTCTT
ATAATTTATTAAGAGTGAATCAACCCCGTAAGAAAATTTCATTGTTCTAA
CCAATTTTTTAATATATAAGTAGTTTTTTAACATCAACCCAATCTCTAGA
AAGACCATTATAATTTCCCTTTAATTATAGACAACTTTATCATATAAAAG
TTTTTTTAAATAAATCCTCTTATTGTGACTTACACAGACTATTCATGACA
TGCTTGGACTTTCTGGTTTGTCGTGAACATCCTTTTCTTTCTTTCTTTCT
TTTTTAAATTTTACTTTACGTTCTGGGATACATGTGAAGAACATGGAGGT
TTATTACGTAGGTGTACATGTGCCATGGTGGTTTGCTGCACCCATTAACC
CGTCATCTATATTAGGTATTTTTCCTAATGTTATCCCTCCCCTTGCCCCC
CACCTCCTGACAGGCCCTGGTGTGGGACATCCCCTCCCTGTGTCCATGTG
TTCTCAATGTTCACTCCCACTTATGATTGAGAACTGCAGTGTTTGGTTTT
CTGTTC
>Contig23
GCTAAATATAAGCTATGATAAAACAGTTGGCCCTCTGTATCATGGGTTTC
ACAACTGTGGATTCAACTAACTGTGGATGAAAAATACTTGGGAAAAAAAG
AATGGCTGCATCTGTACTGCACAAGTGCGTGCTTTTATTCTCGTCATTAT
TCCCTAAGCAATACAATATAACAACTATTTATATAGCATTTACGCTGTAT
TAGGTATTATAAGTAATCTAGAGATGATTTGAAGTATACAGGAGGATGTG
CTTAGGTTACATGCAAATATTATGCCACTTTATATAAGGCCCTTGAGCCT
CCTCAGATTTTGGTATCCATGGCAGTCCTGGAGTCAATTCTCCTGCAACA
TCTCCATTTGTTCAGATTCTCTTCTATATCATGTTTATATCAGAAAATCT
ACATAAGATTTTTTAATGTGTTCATATAGGTTTTGTGTATTTTTGGTTGT
TAATCCCTAGATATATGCAGTATTTATTGCTATTATGAGTAGTGTTTCTT
TACCATGTATTCTAGTTGGTTATTGCTGACAGAGAAATGTTGCTGGTGTT
TCTAAGTTACCTTGTTTCTAACAACCTTGCTGAACTCTTATTAGTTCTCA
TAGTTTTAATTAATCTTTCTTAGTTCTGATAACATAATCTGCAAATAAT
GACAATTTTATATCTTTCTTTCCAATGCTTATATCTCTCAGTCCTCTTTA
TCCCAAAGTATTTTCCAGGATCTCCACTATAACATTAAATAGTAATAAGA
ATTTCTGTCTTGTTACTGATCTTAAGGAGAATAAATTTAAATTTCCTCTG

FIG. 4 (6 of 61)

TCAGGTTTTATGCTTGATATAGATTTGTGATATATAGCCTTTCACAGGTT
AAAAAAAAATGCTTTCCTAGTAGTCCTAATTTTTTAAAAAAATCATCATA
AATAGATGTTGAACATTATCAAATGCTTTTTCTGCATCTATAGAGATAAT
CATATGGTTTTTTACTATTTATTAATGTAATGAATTAGACCAATTTTCTA
ATGCCAACTCTTTCTTGTATTTGTAGGGTAAATCCTATGGGATCATAAAA
TACTTTTAATACATTGTTAGATTTGAAGAGTTAACGCCTTATTTAGAACG
TTTTCAGTCACATCCATAAGTGAAATGGCACTATAGTGTCTATTACTATT
ATATTTTTCTGGTTCTGAAACCAAAATTATACTCACCTCATACAGTAAGT
TGGGCAACTTTTGTTCTTTTTTTCTGAAACAATTTGTGTATAGAAGAAAT
TAACTGTTCCTTGAAAGTTTGATAATAATCATCCAGAAAATTATCCCCAT
CTAGGGCTTTTACAAAAAGGAGACTCTAGAATGCCATTTCGGTTTCCTTG
ATGTGTATTGGCCTCTTTCATTTAGGCTTTTGGATTTTTTAGGGCATTTT
TTCACTATAGGCTTTTTACCGG
>Contig24
CATAAACTTCAGGTTGGATGTTCGGTCAAAGTGGTCCGGCGATGCGAAAA
CGAGAGGGCTCGAGGACTGGGCAGAGAACTATTTGAAGGTATCTCTCAGG
GGAAACCAAGCGGAAGGCGGGGAGTAAAATTGGAGGGAGCGACGGCCTT
CAAAGAAGGGGCTTGCATTAGATCGGCGAGATCCGGGAGGGTCTGGTGGG
GAGAAATGACTAGAGGACAAATCTAATGGAGAGACAGACGGAGATAGATA
TCGTGACAGAGAGAGGGACAGTGACAGCGCACAACAGTGCAGGGTCCATG
AGTACAAGGCCCTTAAGTGTACACCCCAGCCGGAGTCATGGCAATTCGAT
TCCTGTACTGACCACCCAGGATTTGGGTAGACTGTACGAGTTAATGAGCA
TGGTCCCCAACAAGACTGCTTCGACCTCAGATGCAAAGCACACTTCAGGG
GTCCCCAAGCCACTCATGTTTTTTGAATGACTGCCATAAGTTCAAAAATT
CCCACAATTCTCTCAGATTCAATAACTGGGTATAACCACTCATAGAACTC
AAGAAAATGCTATCATTATTATTACAATTTTATTATAAAGGATACAAATC
AGAAGGACTAGCCAAATGAGGAGACACATAGAGAGAGGACTAGTAAAAAA
CAGAGCTTCTGCGTCCTACCTTCAAGGAATCAGGATGCACCACCCTCCCA
GCACATCAAGTGCTCATCAACCAGGAAGTTCCTCTGAGCTCCAATGTCCA
GAGATTTTAGGGAGGATTCATTACATAGGTATCATTGATTAAATCATTGG
CCATGTACTTGAACTCAATCTCCAGTGTCCCTCTTCTCCCTAGAGGTCTG
AAGGGTTGGCTAATATCATGTGGCTCAAAGCCCCAACTCTAATTACCTTT
TTGGTCTTTTCAGGGACTAGACCCCATCCTGAAGCTATCTACAGGCCCTG
CCATGAGTTAGCTCATTAACATAACAAAGACACTTATATTACTCAGAAAA
TTCCAACAGTTTTAGAAGCTCCATGTCAGGAACCTGGGACATAGATCAAA
TTCTTTTTTTTTTTTTTTTTTTTGGAGACAGGGTCTTGCTGTGTTGCCCAG
GCTAGAGTGCAACGACAGATCACAGCTCAATGCAGCTTCAACTTCCCAGG
CTTAAGTGACCTTTCCACCTTAACCTTCCAAGTATCTGGGACCACAGAAA
ATGGCTAATTATCCTGGCTGATTTTTAAACTTTTTTTTTTTTGTAGGGATG
GGATCGCCCTGTGTTGCCAAGGTTGGTCTCAAACTCCTGGGTTCAAGCAA
TCATTCTGCCCTGGCCTCTGTGATGGTTAATACTGAGTGTCAACTTGATT
GGATTGAAGGATACAAAATAATATTTTTGGGTGTGTCTGTGAAGGTTTCG
CCAAAAGACATTACTTTGAGTCAGTGGACGGGGAAATCCCCCCTTCCCCA
TGGGACGGGGAGACCCCCCTCCATCCAGGTAAAAAAATCTAATCACCTGC
AATGTGGCAGAAATAAAGGAGGGAAAAAACGGGGACCCCTANATGGGTTA
TTCTCCACCTAATTCTTCCCCCAGG
>Contig25
CCATGTATTTCATTTCTACAGACCCTGAGATGAATTTGTCATTGCCACGG
GGTCCTGAAGTTCAAATACTCTATTTGGTATCCTGCCCCTGTGGTTAACT
GTGATCATTTCACTCACCTTGTTTATGATGAGAGGTGCCACCATCTGGCC
TCCTCCACTCTGCAATCCTGTTAATTCCTATCAAAGCTGAAAACCTGCTG
CAGCACCCACACCATCACCTCCAGCCTAGAGAGGGAAGCTACCAGTGAGC
TCTCCTGGATGCCGGTGTGCCCCTCGCCAATACATTTCTTCTTAGTCCCT
TGGTCATCCTGAGGTGTGTGATTAATGGACAGCTATGTGGATTGCACATA
ATAGATGTACTCCAGCATCTTCATCCCTGATTTTCCTTTACAGAAATCAC
TCAACCTTAGCAACATGTGAAAATCACCTAAGGACATTCTTTAAATCCCT
CTGTCCACATGGCAACACAAACCACTTAAATAAGAATCTCCAGGGAGTCA
CTCAAGCATCAATGTTTTTTAAAGCTCCAATTTTAAGGATCATTACATTA
TGTCGAAGAAATTATAGTATTTCAGCCTTACTGACTGTAAACCACCACCA
TATCTAAGCATCCATTAGTCAACCTAGCAGACAATAAACTAACATTACCT

FIG. 4 (7 of 61)

CCAGGTACTCAAATCAATTCATTGCATCCCAAATCCCAGATGGGCCCACC
CTTATTGACAAATTCAGCCCAATCTTGGTTGAACACATTTAGAATATATT
TCCATGAACAATATCCGGTTGACGAGTTTCTTTAACTTTTTGGAGTTTAA
GCCATTTCCTTTCACAGTAGCCTTGTTAATTCCCTGTCAATGCTCCATGG
GGGTCATGAAGAGACCTCTTATTAACTGTGAAGCAACTTGGCTCAGGTGC
AGACACTCAAATGCTTCACATGCAGTGGGAAAAGAGAGTGATTGTCTAC
>Contig26
TTTAAAAAGAACTGAGTCTTTATTCAGTCGATTCTTCTAATCTATGAACA
TAGCATCTCTCTCAAAGCATTTAGTCCTTCTTTAATTTCTGTCATTAATT
TTTTAAAATTTTCATCCTAAAGATTCTGTATATGTTTTGTTGAATTTATG
CTTAAGCATTTCACTTTCTTGGTAACAATTATAAATGATTTTGTGTTTTT
TATTCCACTAGTTCATTTTCAGTGTGTAGAAAAGCAATGAATTTTTGTGT
GTTGATCTTTGTTCCAACATCTTGCAACATTATTGAACTCATTTATTAGT
TCTAGGAGGTTTTTTCATTTTTCTTGTAGATACCTTGAGATTTTCTATAT
AGACAGTCATGTTGTCTGCAAACAGGCACAGTTTTATTTCTTCCTTTTCA
ATCTATATGCCTTTTTTTTTTTTTTTGCCTTATTGCAGTGGCTAGAACTT
CTAGCACTATGTCAAATAGCATTGGTGAAAGCAGACATCCTTGTTCCTTG
TCTTAGAGGAACATTTGGTCTTTAATCTTGATTTAAAAAATTCCTTGCAC
TAAGTTACCGTGTTTTGCGGGAGGGAGAGGTGGGGTGAGGTGGGGATTTC
CCCTAATGTTTACAAGCTGGGATTTTCTTTTTCCTGTGTCTAATTATTTT
CCTCATTGGCTTGAAAAATCTGATAAAACATTTTAGGACTGTGTATAAAA
TAGAATTAGCCAAGTGCAATGTCTTTATTCAGAAGAAATTTCATGGACGT
TGTGCCTACTCTCTTGGCTTCCTGGCTTCATGGCTTTCCAGATCCCACAG
TAAGCTCTGGATAGTAGAAGTTATAGTAAGACTGACTTCTAAATAAATGA
AGTGACTTTAACCTTACTGATATGGCTTAAAGAAAAGGAGTGGCCTTTAA
GATCCATGAACTTCTCAAACAAAAGTGATAACGTTATCTCCATGCATATA
TAATACTAAATATAATGCAACTGAGAGAAGTAGGCTGTGGTAAGAAAGGA
GACCCAAGTGCCATCTGAAGGCAGCACTTACCACTCTGCTTCATCCCACC
GAGGAAACAAAGCATGAGTATTGCCAGATTTTCTTCTGTTTCAAGAAAAG
CCAGAAATCCAGGTTTTTGCGTGAAATGTCCTGATTTTAATGTTGGGAAC
TAATTTATATTTTGAAATAACATTGTGTGGGACAAGTGAACTTGTATGTG
GAACTGCTTTCTCCCAGTGGCGACCAGTTTGGACCGTTGATACTCAGCAA
GTTCAGCCAAGTGCGCCTTGTCATTGTCAGTCATCAAGGTGATGTGTGAT
TGGTCAAGCAATTAATTTTGCTCAGCATCTCGTGTGTTTTCAAAAGAACT
GAAGGTTCATTTGC
>Contig27
TTTCAGAGCACAATGCGTATTCATAGTATATTGACTTAATTTCTAAGTGT
AAGTGAATTAATCATCTGAATTTTTTATTTTCAGATAGGCTTAACAAATA
GAACATTCTGTATATAAATGTGTAAATTAGAGTTAATCTTTCCAATCACA
TAATTCGTTTTATGTGAAAAAGGAATGAACTGTTCCATGCTGGTGGAAAG
ATAGAGATTATTTTTAGAGGTTTGTCGTTGTGTTTTGGGATTCTGTTTTC
TTTTAAAATTGTAAATATGTACTTGTGTGAATGATTTTTTAAAATGATTT
TACCATTTTTGGAAGGGTATTTAATGATAGAATATCATCGAGCCAACATG
CACTGACATAGAAAGATGTCAAAGATATATTAAGTGTAAAATGCAAGAGG
GAAAACACTATGTACAGTCTGAGCCAAATCAAAGCATGTATGTTTTTTAT
ATGTGTACAACAAAAGGTTTGGAAAGATATGCGCCGAATTGTTAAATGTG
GTTTCACTTGAGGGGGTGGGAGGATGGGGCCCCAGAGGGGTTTTTATGGG
GGCCTTTCACTTGGTATTTTTTTCATTTTGTTCTGTTTGAAATTTTGTTT
TTTCTTTTTAAATGGAGTTTCACTCTTGTCGCCTAGGCTGCAATGTAGTG
GCGTGAACTCAGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGTGATTCT
CCTGCCTCAGCCTCCCATGCCTCCTGTGTAGCTGGGATTACAGGCACCCA
TCACCATGCCTGGCTAATTTTTGTATTTTCAGTAGAGATGGGGTTTCACC
ATGTTGGCCAGGCTGGTCTGTAATTCCTGACCTCAAGTGATCCACCCACC
TTGGCCTCCCAAAGTGCTGGGATTTCAGGTGTGAGCCACCACGCCCAGCC
CTGTTTAAATTTTTTATAAGTATGTACTACTTTTGTAATCAGAATTATTA
GAAAGCATTTTACTGATTTAAAAGCTTAGACATGTTCAAATGCCTGCAAA
ACTACTTAACACTCAGCTTTAGTTTTTCTAATCCAAAAAGGCCGGGCAGT
TAATCTTTTTGGTGCCAATGTGAAATTTAAACGGTTTTATGTTTTTCCTG
TGTTGTGAATGAAAAATATTTCTGAGTGGTGGTTTTTTGACAGGTAGACC
ATGTCTTGTCTTGTTTCAAAATAAGTATTTCTGATTTTGTAAAATGAAAT

FIG. 4 (8 of 61)

ATACAATATGTCACAGATCTTCCAATTAAGTAGTAAGGGTTTATCCTTAA
TCCTTGCTAATTTAAGCTTGCATAAGTCACTTTACTAAAAGATCTTTGTT
AAGCTAGTATTTTAAACATCTGTCAGCTTATGTAGGTAAAAGTAGAAGCA
TGTTTGTACACTGTTGTAGTTATAGTGACAGCTTTCCATGTTGAGGTTCT
CATATCACCTTGTATCTTGAAGTTTCATGTGAGTTTTTACCATTAGGATG
ATTAAGATGTATATAGGACAAAATATTAAGTCTTTCCTTTACCTAAGTTT
GCTTTCTTGACTAGTAATAGTAGTAGATATTTCTGTAATAAATGTTCTCT
CAAGATCCTTAAAATCTCTTGGAAATTATAAAATTATTGGAAAGACAAGA
ACAGTTTTTATTCATTATATGCATTATTATCG
>Contig28
CTTTCTCAAGAAAAGGGAACTGGAGCAATTAAACATATGTAATTTTTTTT
TAAAAAACCCTAAACCTAAACATCTACCTATATACAAAAATTAATTAACA
ATGGATCATGGACTCCAATGTAAAACATGAAACTCTAAACTTCTAGAAAA
AAAACTGGAGAAAACCTTTGGTACCTATGACAAGGCACAGTTTTTAGACT
TAACACTAGAAGTGTGAACTATACAAGAAAAAATTAATAATTTGAACCTT
ATGAAAATCAAATTATTTGCTCTCCAAAAGACCCTGTTAAGAGGATGAAA
ACTAAATTACAGATTGAGAGAAAATATTTGTAAATCACATATTTGACAAT
GGACTTGTATCTAAAATATCTAAAGAACTCTCAAAACTCAACATTAAAAA
AAATATCTAATTAGAAAATGAGTGAACATTTTACGAAAGGGGCCTTATAG
ATTAGCAAATAAAACACTTGAAAAGATACTCAGCATCACTAGCCATTAGA
AAAATGCATATTAAAACCACAATAATGTATCGCTACACACATATAAGAAT
GGTTTATGAAAAAATAGTGATGACACCAACTGTTAGTGAAGATGTGGAGA
AACACTCATACATTGCTGGTAGAAATGTAAAATGGCATAGCCACTGTGGA
AAATTATTTGGCAGTTCCTTTTAAAACTAAAAATCAATCTACCACACAAC
CCAGCAATTTCATTACAGGGCATATATCCCAGAGAAATGAAGATTTATGA
TCACACAAAAATCTGTACACAAATGTTTTATGGTCACTTTATTCATAATA
GCCAAAACCTGGAAACTATCCAAATGTCCTTCAATGGGCAAAGGATTAAA
CACACTGTGATACATCCATACCATGGAATACTACTCAGCAATAATAAGGA
AAGAATTACTGCTACACACAAGTTGGATTAAACTCAAGGAAATTGTGCTG
AGTGAAAAATTAACAAGCCAATCTCAAAGGACACATACTTCATGATTCCA
TTTGTATAACATTAATTAACACAATTAATTACAGAGATGGAGAACAGAAT
AGTGGTTGCCAGGGATTATACATGGTGGACGCGGTGAGGCGGGCCTCCAC
GCCTTGGAGATGAAGGGGGCTACACCCTTTAAAGCACACCCACGAGAGAG
TTTTGTGCGGAGGGGCCCAATTTAAGTACTCCGCCCCGGGGGGGGAACAC
AGGGGCAAACAAAAAAAAATTGGCCTTGGGGGTGACCAAACACACAAAAAA
AAAACAAACACACAAAAAAAACAACNATGGGTGGAGGATTAATCGCCAAA
TCTGAGTAAGCTATCTGGACAGTACCAATATCGATTTCCCAGTTTTGATG
TTGTACTATAATAATGCAAGATGTTAACATTGGAAGAAGCTGGCTGAAGG
GGGCTCAGGAACTCTCTGGACATTTCTTTGTACCTTCCTGTGAATCCATC
ATTATTACAAAATAGGACATTTTCTAAAGGTTAAATCATTTTAATTTTAA
AATGTCCCTGTTACTGTTGAAACTCACATCTCCATATACTGATCAAGAAC
AGCACTAATGGCCCCTGGCCTCCAGGAATTCACAATTCCTACTGACTTTT
CTTTGAAACCTTGGCCAAGTCGCTTCTCTTCTCTGGTCCTCAATTTTTCA
TCTTCAAAATGAAGATTGAATGACTATTAAAATCTCTTGCAATTCTTGAG
ATGAAGGGTCCTAAAGGAACTGAAGAGGATGCCATGTAATGTAAATATGG
GTTTTTACTCCATCAGCCAGCCAAGACAGAGGGCAGACACCAAGACATGG
TAACCAAGGAGGCCATGTGTAAACAAAGACCATTTAGACTTATGCTCTGG
CCTTTGCAGCCCAACTGGTGTGGCCAGTTGGTGGGGTATGAAGAAAATGG
GGCCTTCCAGGAACCATGTTGAGTGGAGATAAGCAGGGAGGAATGCAGAA
GACATGGGGGCAGTGCCAGTCTCAGCCCGAGCCAGCTACACCCACACATG
GTTATGAAAGACTGACAGCCTGTAAGNTGAACACAGCCCTGCCTCTCTTA
GATAGGC
>Contig29
GCAAATATGATCTCAGATGTGGATTTACTGTAAAGTTCATCAAATTTAAA
TTTCAGAACACTTAATCTGCAAGAGTCCTTTCCAAGACCCTATACCTAAT
TTTGTGTTTACAATTTTATATTTGTTTTCTTAAAGAAGACCACCAATATA
AACTATATCCAGCCTTCATGATAAGTACATAGGAAACTATGCAAATAAGG
GGGAAAAAAAAACAAAGAAAAATACCTAGTTTACTAATGGTTCACTTCTGA
ATAGCACATATTCATAATGATACAAGCACTCATTACTAGTCTAGGAAAAT
GAAGATATAATTGCATTAGGAAGATCAAGAGGTAGGAAATGTGGATGTGT

FIG. 4 (9 of 61)

GTGGTATAGACTAGGGCAGGACAAAGAACCTAAATCCTCATTTTCTAAAG
ATAATTGTTAATACGTAAAACTCAAAATTCAAGAAGTAACAGTAAAAGCG
GTCATTAAGAAACAAGCACTAAACACCAGATAGGAAGCGAGAGATGGGGG
AAGAGGGCGACAATCTGATTATTTTTTGCAACAAATTTTGTAAAACCATT
TGACTGTTTACATGTAGAACTTGGATCTTTTTTAAAAAACACAAAATAAT
AATACTATTATTTTTTAACTGGATTTTTGAAAAAGAAGATAAAAGTCTCA
TTTTAGTAATTAAAACTCATTCCAGGTTAGTCCACTCAAAACTTATATTC
GAAAATTAAAACTTTGGGAGGCTGAGGCAGGCAGATCACCTGAGGTTGGG
AGTTCGAGACCAGCCTGACCAACACGGAGAAACCCCGTCTCTACTAAAAA
TACAAAATTAGCTGGGCGTTGTGCATGCCTGTAATCCCAGCTACTCGGGA
GGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAG
CCGAGATCACACCATTGCACTCCAGCCTGGGCAACAAGAGTGAAACTCCA
TCTCAAAAAAAAAAAAAAAAAAAAAATTAAAACCTCTGGAAGTTGAGTTTG
CAAATATTCATTATGCTCATTTTTAACTTGTATGTTTGGAAAATGTCATG
ATGAAAATTGAGGTTGGGGGATGAGAAAAAAAGAAAAACATCAACCCCAC
AGCCCATTCAATTTTCAGCCCGACCCACAGCTCCGGGGAAGGGCAGCAGG
TCCATCCTTCACTCTTTCTTCACCTCTTTCCCCTCCTTCTGGCTCTTCCA
CCTCTAATTTGGAGCCCAAAAAAAGGCACTGGGAAATGGAAAAGTCTTTT
GTACGTGGTACTTGCCGGGGAAGCTGCCATGAAAACCTGGCCCCACGGTG
GGGAGGGAATGCCCANCTGAGGCCTCGTGCCCATGCTAGGATAGACTCGT
CCAAACATGTCAGGTGGTCTGACAGGGCAAGCANCANGAAATCATGTATG
AGTATGAACTGATCTGTATGCAAGGGCGGGGAGAACACGCGGAGGAATGG
GGCGTGAGAAAACAGCACAGTACGTTTCTTTAGCAGCTGTCTCTGCTCAG
CCATGGGAGGTCACAGAGAAAGAGGCTTGGAGGCGTTATTTTCACTGTGA
GATGTGAGTGTAAAAAAGTGCCCAAGACACAGTGAGTACCAGGGAGATGC
CCTCTTTCCTACCCGAATGCAGAATGGCCACAGGCCTTAAAACACACACA
TGGGTCCTCAGAGGAGAGAGGCCTCCACAGTGGACACCCGCATTCTCCCC
TGGTCAGCAGCAGCAGGGCGAGTGCTGGGCCATCATGAAGCTTCACAGGC
AATGAGCTCTCAGCAATAACAGGAACAGTGCCTGGGGGACTGTAGCTGCA
AGACCGATTTTCATGTAAGATGGCCTCTGAGGACTCCGAGATACACCAGG
CTGAGACTAGCTGGCAGCTCCAAGTTGTTGGTCAGAAGAGAACAGGAACT
AGGGAAATTGGAATTACTGTTACTACAATTCCTTTACATCCGCACAACCA
TGAGGTCCAGCGATTTTCTATTATTTTTTTTTTAAGACAGGGTCTCAGT
ATGTCGCCCAGCATAGAGTGCATTGATGTGATCATGGTTCAGTACAGTAT
TCACGTCCCAGGCTCAAGTGACCCTCCTGCCTCAGCCTCTCAAGTGGCTG
GGACAGCAGTTGCATGCTACCAGGCCAGGCTTTTTTTTTTTTTTTTTTA
GTTTCTGTAGAGCACATAGC
>Contig30
GGTTAACAATGGCACAGGGAAACAAACAGTTCCAGGTGCAGGGGCTCTAA
ATCTATCATAAGATGTTAGGTATGGGGGCTCTGCCGGACACAAACTCAAG
GCTTTATGCTGTTATCTCTTGAGCGAAATCCTGGGAACTTCGTACATTGC
TTGCTTCAGTACCTTATCAGTTAATCGGACTCTTTGATATGTTGGGAGTC
AGCGTACACAAGTTAACTCCTTGAGGAAGGGGGTGGTAAGGAGTCCTTG
ATGTCTGGTAAATGAAGGAGCGAAATCGAGTTCCTCTGGCTTTCTCAGCT
AAGGGAGAGCTTATTCATGTGGAAACAAGGCTAAGTGATTAAGGGAGAAA
GGGAGAGTCTGAAAACAAGGTTAGGTATTACAATGTCAATAAAATTGGTC
TCCTTATACAGTCCTATGGTAGATTTCTTTCCATCTTTAATCTCCCTCTA
GCACCACCAGACTTTTTCTCTCTGTACCTTGAGATGTAAATTTTGCTATC
TGAATTTTCGTCTAAGAGTTGTTTCCTTTAATATGCAAATTTAGGGTTAT
TTAGCTGACAACTGCCAAAGTAGTGAAACAAGTTATCAAGAACTTGAACG
TCTAAGGTAGGAAAAAAAAAAGTCTTTATGAATCTATAAGATGTACTTCT
ATTGGCATGCCTAATACGTCTATGTATTTACGTGTTGTGTACACAGTTTT
TCACTACTGAAAATATATAGAGGAGTTCTAATTAATTGACTTAAGACAAT
AAAAGCGCTTGAATCAAATACCTTATCAGGAAAAAGGAAAAGACAAGTCA
AATGCTTGTTCAAGTTTATATAACTTAAGTAAAATCTTTAATAAATAAGC
TAGCTTTAACATTATTTGAAATGTCTTAAGAATTGCCAGCAGGTTCTGGG
TTACAGAACTAGTGGGGGTGCAGTGGGGTGAGGGTTGGTGGGGTGGGGGG
TGGTACGGGGGCTTTGTTTTTTCTTGCTGCCCCCTTCTGGGTTGGGGAAG
TGGCAGGACCTTGGCAGCACCCCGAGCCGGCATGGCGTTAATAATGGAGG
GATGCCAGACCCAAGTGGCTAAGGCCCGGCTGCAGAGCCAAGTTGGCATT

FIG. 4 (10 of 61)

TCCAGACTGGGGCTCGGGCCGCACCCTCTCCAGGACCCTCCCCTTGTACC
GAGCAGATTGTCGCGGGCAGTTTGGGCCAGCTGTCCTGGCGTGGAATTTC
CCAAATTCAACAAATCCTCCAAGAAATCAATCCATCCATTCATCCATCCA
TCCATCCATCCATCCATCCATCCATCCATCCGTGGCAGATTATGAAGCAT
GGATCATTACTTTGGGATGTGGATATATTCAGTTAACAAGGAGCAGCTT
TCAAGAGCTGGATTTTATGCTTTGGGTGAAGTTTAGAAACACTAGCTCCC
AC

>Contig31
ACCTCATGTGCTCTAGCGCCTCTTACCTCATGCCCTCCACTCTCAGTCTT
GCACTCACCCTGCCACACTCAAGGGCTTCCCCAGGTTCCTTCTTAGATTC
CACCGATAGCTCAGGGACTTTGCACATGCTACGGTCTCTGCCTGGCTCCT
CCCCAGATCTTCTCATGCCTAGCTGCTTCTCATCAGCACCCCTCAGAGAC
TGTCCCTGCCCCACCTCTCCAGGTTCCATACCTGCCACCCTCCCCCAATC
ACGTAACAGTTTCTTCACAGAGCGAGTTACCATCCCAGTATTTCCCTAAC
TTATTTTTGTGACTGGTCTGTTGCCTGTCTCCACCACAAGAACATAAGC
TGCATGTGAACAGGAGCCTTGTCTATCTTGTCACCCCAGTGGCTGTGACA
TAACCTGATACACATTAGATGCTCAATGATGTTTGATGAATGAAGTGCTG
GTAGTCCAACTGTGTTTCCTTGTCTGTGTAAGTATGTCTGTTGTGGTTTC
CTAAGAACCTACAGCTCTCCCACTGTGACTCCTGTTCTATGGTCCTGATT
TGCTGGACTAGAATCCTAACCTACATGCTTACTCTTAGTGTCCTCCCCCA
GAGGCTGAATCCCAGTCCCTAAACCTCCACCAAATGGCTAAGACCTAGCT
TCCAACCAGACAGGCCTACGCTGAGACCTCAGCACCGCCCTTCTGCGGTC
TCATCCTTAACGCATCCTTCAGGGCCCAGCTTAAATGTCTCTTCTCCAAG
GAAGGCTATCCTCTTTCTGCCCCTCAGTGCTCTCCATGCCTCCTCTATGC
CTCCATGCCTGCTTTCCAACCCTGCAGAGGTGGAGAAGTTGCTAATCTGC
TGTGTTGACATGTGCTGGGGTGCCTTGGGCCAGGGAGCAGGCTGGTGGTG
TGCTGATAGCCCGTGGCTGTGCCCAGGTCCATGCTCACTTCCTGAGCCCC
AGTGGAGTAGGCTCCCTTTCCCTTATTGCAGCACTCAGAGGAAGGACGTG
CTTCTTAGGACAGATCTGGCCAACCTCTCCCTCGTGAGAGAAGGCCCAGC
CATCCTCTTGCCCTCTTTCTTTCTCCTGCCCCCGAGTAATAAAGGTGCCT
GGTCAGAGCCTTCTAGAAGGAGACCCAAACATCCACCACACATTCCCAGT
TCCAACCGTCATCCACATGGCTGGCTGTGCAGGTAAACGCAGAGTCTGTT
TCACACACCCAACCATCTAGTATTGGATGGGAGGACAGTAGCGTGACACT
CTTCTCCAGCCTTGAGCCCTACTGTGGGCCCCACCCAACCCAGATACCAG
AGGAGCCCTGTACTGGGATGCTATTGGATGCTTGTCCAGTCATGTACAAA
GTTAGCCCTTTGTTATATAGAGTTAGCTACGTACATCTTCCTCTGTAGGG
AACCCAAGAGGGGAGAAGAGATATGTAGTAGGATTTAACCTGCAAATCCT
CTGCTGAGCACCGTGCACTACATACAGTGGGTAGCATGTGGTAGGTGCTC
AATAACTATTGACCGATCTATTGAATACACGTAAGATCGTGACACTATCT
AAAACGNGGGGTGTGGGGGAAAAACCCCCCCCTTGTTTAGGAAACCCAAA
TTGGACCGTGTTGGC

>Contig32
GCGCGATTGTGCTAAAGATCATGCATGCCTGATCAAACGTCCCCATATGG
CGTCTCAGAGTCAACTCCTTCCCCATCAGTGCCCTGACTTCGGCATAACA
AACCTGGCAGGTTAAGTGATTAATCGGTCCTGTACAACTGTAGCCCTTAG
CAGGAAGCACTAAGCTTCGTTTTCATTTATTTCTTCCCTGGAACTGCAAG
AAATGAGGGATGCCTTCCGCCATGAAGTTTTGCTGATTGTCCACTTTGTT
CTCAAGGAGATATTCACAGTTTTTAATTTGTCTTTCTCTCCTGCATGGTC
TCCAAACCTGTCCAAAGAAGCCAGCTGGCTCCATCATCTGTAAAATCACC
ATTGTCACCAGAGCACTTGACTTCCTGTTGCCCTACAATCCACCTGCACT
TTATTTCCTGCCACCATGATAATGTAGTGTTACTACATTTTACATTCAGC
TGTAAGAAATGTTACATTCATTTACTTAAATCAAATTAAGTCTGCTCACT
CAGTCCCCCACAGTGACCAACTTATAAAAGAGAAGGTACATTTCAGTCAT
CACTGAGGTTCTCTCTACCACTGGAAAACTGAGGAAGGGTCTGGAGTCCA
CAGTGGTTAACATCATTGCCTCTGTTTTTTCTCCTACTCAATGTAACCAT
CCAAGGTTACTCACAAATTCACAAAAAGAGGTCTTCACCTCTGCTCTCAA
GACCCAGAGGGCTGGGTTCTAAACTCAAAGGCCAATGTTCCCCAACTTTT
TGCATTGTTTCAACATTGGGGAAAACTCGAGGGGATTCAAGAATGGTTAT
ATAAGTTTTGTGGAAAAATGTATAATTTTTTAAAATTAAAATACAAAGTA
TTATGGAAAGCACTAAATATTGAATTTATATAAATATTCCAAATATTTTT

FIG. 4 (11 of 61)

CTAAATTTTTAGTGAGAAACTTGAGCTTGCTTCTGTGAGATATTTATTTT
AAAACAGATTTGACACTTAAAATGTCTAATCAAGCCTTTTAAACCATGAT
CTATCTCTTCAAATTCTTCAGATGCCACCATCAATAAAGAAACTTTGTTC
ACACAAGTAAGTGGTAGCAAATGGCAGGGTGTTTATCATTTTTTTTTTTT
CTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG
CGCGATCTCAGCTCACTGCAAGTTCCACCTGCTGGGTTCACGCCCTTCTC
CTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCACCTGCCACCACGCC
CGGCTAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCC
AGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCCTCCCA
AAGTGCTGGGATGACAGGCGTGAGCCACCGCGCCCCCGCCGTGTTTATCA
TTTTTTGCCTGATGAAATTTTCCTTGCCACTACTCTGGATGGTTTGATAC
ATTTAAATTGTGCTTCCAGGGTACAATTATCCTTTAAATCTATACCTCTT
TCCTTTCTTTTATTGACAAATATAATGTTACACTTTTCTGTCATTGCAGC
CACACCACCAGTACACAGATCCCAACAGAGTTGTAATATTTTATTAGTTT
CAGAGTTTCAATATTTTATCACTTTCAATACTTCATGTGCAGGAGTTTTA
TTTGGTACTTCTTTACAAAATAAATGATGTGCTTCCAAGCATTTCTTTTC
AATAATTCCAATCAATGTTATTAACTGAGTAATACTAGTATCTGTTTATT
CATAAATTCACAGGAAATGCTTTTTTACTTATTAGTCTTTGGAATTCTGT
TGTTTGTATAAACATCTTTCATGATGGCTTTGTGTCTACCAATAGCACTA
TTGCCAAAAGGCACCTTTTTCTTGTTCCTTTACTTCACTGGTCCGAAGCC
TGGTACCAACAACTACCACACAGACTGGGAAATGAGCAATTTTGCCACGT
GCCCTTAGCTATTAATGGTGGCACTCCATAACTAGCATCTTAAGCTCAAT
TTCATGAAAGAAATGTGTTTCTTATTTTGTACTTGCAGGCACTTTTTAAA
CTTGTAATCTTTTATTCATACTTTAAAATTAAAACAGAGTAATAGAACCC
ATAGAAGGAAATCAATACCCACGAGTCCATACTGATATAAATAAATAGTT
ACATAAATAAATGGGGGGAGAAATAACAGCTCTTCCTTACAGAAAAAATTT
CAATTAATAAATGAAGAAGGAATTAGGGAAATACAACGTTACCATTAAGC
AACCACAGTAATAATCATTACAGGCAATATCCAAAAATAAATTCCAAAGC
CAGTGGGCAAAAGTTTGAGGAGATACAGGATATTAACATAGTCTCCAAAT
AGCTCATGCTATTTATAAATTACAAAAGGAAACATAACAACTGTATAGTG
AAGAAACTCAGCAGACACCACCTTAGCCAAGTGATCAAGGTTAACGTCAC
TAGTAATAGGGCTTGTTGACATACTGGACTCCAATCTGATACACTGATAA
GGACACATGACTTCTGCAGTATTCTTACCAAAAACAGAATTCTAATGTAA
TTAAGGAAAATGTCAGACAAACCTATTCTGAGAAACATTCTATAAAACAA
CTAACCAATACTTTCAAAATTGTCAAGGTCATAAAGACCAGGCGATGGTC
ACAGATTTGAGGAGACTAAGGAGATACAACAACTAAATACACAAATGGAA
CCATGGCATTCTTGATTGGATCTTGAAACAGAAAAAGGATATTAGGAAGA
AAAGCTGATGAAATTCTAATACATTCTGTAGTTTAATTAATAGTATTGTA
CCAATATTAATTTCCTAGATTTGATCATTATACTATGGTTAAGTTTTTAA
CATTAGAGGAATCTGGGAGAATGGTATATATGAACTCCACTGTTCATTCA
ACTTTTTCAGTAACTATTATTTCAAAATAAAGTT
>Contig33
GGGAGCGGCGGCCCACGCTGATCTCTAAAGCTTTAGACCACATTGGCTCG
AGCATGGTCATGGCCGTTTCCTG
>Contig34
GACGTCTTAGCGCTATATTATAAAGAAATATTCACCTCCCTGCTGAGCTT
ACAGGGTGTACCTAATGTCCAACAATATGAAATCTCTTCAATGAATTGCA
GCACGTCCATATATAACCCACATGGAAGCTGTCCTCTTTCCTCACCTTCG
AACTTCCCATGCCAAAGAGGGACCTCTTGGACTCAAATACATCTTAGCAA
TATAGAAGATGCTGGAGACTTGTAGGAGAAGTGGAGAGGGTTTACAGTGT
AGCCCCACAGAAAACAACTTATGACCCCATCAGTCACTTGTCCCTTTTTT
CCATGCCTCAGTCTAGTCAGGAAACCACTAGATCCTGGATGGCTTCTTCT
CCCTTCCCCTCCTTTCTCTTCTCCTCTCCCTCCCTTGCTCCTCCTTCCTC
CATCACCCACTCCTTACTTCCAACCAAAACTTGACTAGCTCCAGTCTCAT
CCCTCCTTATTGAAAACTATTTTACTCAGCCCTCCTCCCCCACTCCTGCC
CAATCTTTATTCCTTACCTACATCAGACTTCACCAAAACAAAGGCCAGGA
TAATAAACAGGACAAACTCTTTCAAACACATTTTAATGACCATATTTTGT
TATTTTGGTACAATTTGAGGAGTCCCAATCCCCAGGGAAGACTAACAAGA
AGTTCTCCTAACAAAGGTGGGTCTCCCCTTACTAAAAACTCCTGTAATGG
CTGAAAAGAGCATGAGGTTTTCTGCATATCATTACACATTCAATAGAACG

FIG. 4 (12 of 61)

TCATGCAGCTGTTAAAA..GATCTGTAG..AGGCTATCTTGTGACAGAAAG
GCATTGGAGATATACTGTTAGTGACAAAAATAGGTTATAAATGAATTTTT
CCATGCATGCCTCTATATTTATAAATACACACACATAAAAGACAGGAAGG
ACAGACATTAAACATTCATAGTGCTTAAGATGATGCATAGTATAATAGTT
AGGACCATGGCCTTTGGGACAGAAAACTACAGCCTCTCTCCCACTTATCA
GCCATGGGACCTTGGGCAATTTGCTCAGCCTCAAAGCCCCTGTTCCTTTA
TCTGTGTGCTGGGGTTGTTGTAAGAGTTAAGTGCAATACACAGAGAGAGA
GAGAGTACCTAACATGTATTATGTGCTCAGTCAATATGCATCATAGTACT
CATTGTTACATATGTTCCTAAGTGCTTTATACGTTTTTTCCCTAAGTTGA
CCATCTGTTTTTGGCATTATGAAACATAATGATCCTAACAAATTAAAATT
AAAAACATAAAGAATATTTGCCCCAAAAAAATAAAGAACATGAATTCTTC
AAGTAGCCAAGGGGCCATAGACAGAAGTAAGCCCTTGGTGGGGCTTAGTT
GAGAGAAGTCTCCAGAAGGTCTTTCGTGTGTTAAAGAAGAGGGTAACAGG
GAGGAGGTGGGGAGAGATGTTAACTGAGTCTAAATGAGCACCTGGAAGAA
GAGATGGGACAGGCCACTTCTGCCTGGACTCCCTGATTGTTAAGAAGAAT
GAAAAAGAGCAGAAGTCTTCCCTGAGCCCAACTTCACTCCCTGACTTAAC
CTAGTCTTTGCCCCTTCCCTCTCACTCATGGCTACTTTCTGTGGTCACCT
TGTTGTAGAAATGGATGTGCAGCCACCTCATCTTTTTCTACCTCCTTCAC
ATGTTTTAGATAATTTAATGTAGTAGAAGACGGTTACAGCAAAAAATTAC
AAAAATCAAAATATCTCTGCTATCTACTGTTGCATTTCTAACCATCCCAA
AACAGTAGCTGAAAACAGCACTCGTGGTCGAGCGCGGTGACTCATGCCTT
TAATTCCAGATACTCCGGAGGCTGAGGCAAGAGAATCACTTGAACCCGGA
AGGTGGAGGTTGCAGTGACTCAAGATCATGCCACTGCACTCCAGCCTGGG
TGACACAGTGAGACTCCGTCTCAAAAAAAAAAAAAAAAAAAAGCACTCGTG
TATTTTGTTCAAGATCTGTGGTTTGGGCAGGGCAGGGCTCAATGAGGACA
TCTCGTCTCCGTTCCCGCAGTGTCAGGAAGTGTAACTGAGACTGGAGGGT
CACACAGAAGATGGCTCCCTCAAGTGGCCAGCAAATTGGTGCTTACAATT
GACAGGGAGCTGTTGACCAAGGGCCCCAATTCCTCTTCCTATGGCCCCTT
CTCGGGCTGCATGGGCTTCTTTACAGAATGGCAGCTGGATTCCAAGAGCA
AGTATCACAACCTACAGAAGAGTGGAGGAATATTGAAAGTTCACAGTCTC
TTAAGACGTTGGCCCAGAAACTGGCAAAAGCTTCATTTCTGCCATGTTCT
ATTGATCAGTCACAGAACCTGCACCAATTCAAGAGGAGAACATATAGAGG
ACATCTCTCAATGGGATAAGTGTCAACAAATTTGCATCTATCACAATCTG
TCTTTTGGGTACAAACTATTTCTATTCCTCCATTATGCAAAATATACTCA
CAACCTCCCAGGGGTCGCAAAAGCCTCATCCATTTATGGCAAATGTGGCC
CTTTTAATTTATATAAAATAATTTGCGGGGGCTTCCTTTATATTTTTAAC
TCCCCTGC

>Contig35

GTGCAGAGAAGTGATTTAAAGCCCTTCAGAAAGAATGCTTTATTCCCGTG
GAATTTGGTAACTTGCTTGGGTGTGGGGAGGTTTGTCAGCTTTCTCCACT
CAAATTATCAGACCCTTTCCATTTAGTGGTAGACCATTTCCCTCGTCCAG
GCCAAGGGCACATAGTACAGAGAAATAGGGAGTTGTTACCCAGGGAGAGA
ACTTGGCTCTAAACCTGTAATAGAAAGGTCAGTTCTGGTCTGGAGGGTCA
ATTTTGATCTTTGGCTCAGATCCAGGAATTGGAACCAAGGCTTTTGAACA
TTTTAATGCAGGGGATTAAAAAAATGATACGAGTCATTCACGAATATATT
TGCTTAACATCTAAAGAGATCCCTCAAAACACTAGAAAAAATAAGAACAA
AAATCTAATAAAACAAAATTTGTTAAACACATTTACCAAATTTTTTTTTT
TGGTAAAAATTCAAATGTCATAAATAAAGCTAAAGTTCCTCTTGATGACT
CGCTCCTCTGCCCTATTCCACTCCAAGTAACCACTATTATCAGTCTTGCC
AATACCCTTCCAGACCTCTCTACCTCTATATACCATTAGAAGCACATGGT
TTTGCATTGAGGATGTGCAGTGTTTTGTTTTACGTAAATGTTATCACTCT
GTTCTTGTTCCATAATTTGCCTTTTTCTCTCAATGATTTGCTTGGCTATC
TTTCTATTTCAGTAGCATCTCCTTTCTTTTTAACTTACCATTGTTTATTT
AACCTTGCCTCTATCAACAGATATGTAGGTTGTTTCTAGTTGATTTCATT
AAGTATTTATAAACAACGCATCAGTAGATGTCCATAAATTTCTTTACGGA
AGATGGCAAGTAGTGGAATTGCTGAGCCAAAGAACATGTTTAAAAAACCC
AAAAAAACTAGACGCTACCAATTTTCTCTCCAAAATGGCCATACCCACTT
ACCCATACAGAGATGATTTGGAATCTGGCTTCCTCACAAGGTGAGATGCC
TTCACAGTTTCATTCTTCCTGGCATGTCTTCCCTTTTGTATCTGAGAGAG
CTGGCAGAATTGTGTCACTAAATCAAGGATAGAGGGTCAAATGACAGCTC

FIG. 4 (13 of 61)

AAGCTCACAGGCACCTCTGCTTTCTTCC[illegible]AGACCACCTGCTTTCCTGCCA
CCAGCTCTGTTCCATCTTATAGAATGGTTGCCACTTGGGTGTCTGCTCCG
ACAGCCATGTCATCCTTTGCACTGCAGTTATGAAGCAGACAGAGCTAGGA
GAGGGGCTTTGCCAGCCTCTGCCCTAGCTTGGAGAACTTCAAAAAAGGAG
GGTATTGAAGTTGAACTCCCCCAAAAAGGGGTGGTCCCCACACCTCAAAA
AGTGGTGCCTCCGAAAGAAATGTAAAATTCGTGTGGGGGGGAAAAAGGT
TATTTAGAAATTGTTGGCTTGTCGTGCCGAAAGTATGTGTGGTTACGGGG
AGTACGGAAATTTCGAGGGGTGGGGGCGAGGCCGTGTGTCCTTTAGCCCG
GGGTTTTCCCGTCGCATGTTTAAGGGGGGGGAAGAGGGGGGATGTTTTCT
TTCCGCGAAGGTTTTTGAAGAACGGCGTGG

>Contig36

CCCCCCACCGCCACTACTCAACCGGCCGTTCACGAAACAACTCGCCACAT
CCACTAACCCGCTGGCTCACCACCCACCGCCCTCCCGATCCCCCCAATCC
AAACTCAACCCCCACCACCAAGCGCCTCCCCCCTCCCCCACCCTCCAGCT
CAGCCCCAACCTACCACCAACCCCGACTCGCCCACCGAAAACCAACAGCA
AACCCAAATGCCCACAAAACCAGTGTCCAAACCCTCCTTCCCATCAGTTT
GGTGGGCCCATCACCGCTTCCCCTGGCCCAGGCTCTCCTTTTGTGCGCTT
GGAGCAGCAGACTGATCTCCCAGCCTTCACTCACTTCATGTGGTAATCTG
TTGTGTTCATCACTGTCAGAATCTTCTGCATCCCCTCACTACTCTGCTGA
AAACACTCTAGTGGTTCCTCATTGCTCATTAATGAAAGTCTAGATATTAA
ACGTAGAAGGCCCAGCACAATTTGCCCCTATGCCACCTACCTCTCTAATC
TTTTCTCCTTACTCTGACAGACTCTCCGTCTGTCATTTATGTATTCTTTT
ATTGCTCTCTTCTACTTTTAGTATGAACTGGATTTATGGATTTTTTTAAC
ATTGCTTTCAAGTATGGAATAAAGAATTTTATTTATTTATTTATTTATTT
ATTTGAGACTGGGTCTCACTCTGTTGCCCAGGCCAGAATGCAATGGTGCA
GTCATATCTCACTGTAACCTCGAATTCCTAGGCTCAAGCCATCCTCCTGC
CTCAGCCTCCTAAGTAGCTATGACTACGGGTGTGCATCACCACATCTGGC
TAATGGAATAAAATATTACAATGCCTAATCTTAATTTTCAAAATTTTAAA
TTACATTGTACCTAATGCCCATGCATTTACTTTTTTCAGTGGGTCAATAG
CCCTCACTTTGGCAAAGGTCCCAGGCCCAAGGTAAGGCCTTACTTTTTCC
AAACTCATCTTTTGAAAGACATAAGTGCCTGTAAGTTGTACCACATTAGG
TTCTAGGAATTTTTCATCAAAGACTTTATCAGACTATTTTCCTCTAAGTT
GAGAAAGAGCTGGGGGCAGAATATGGCACTGAATGACTGAAGAGAAGGCA
CTGAAATCAGGCCAGAGGTTGCTGGAAAGAGCAATGAGGAACACCAGCAG
CAATGAGGAGCCGGTGATGATTTTGGCTTCACAGGGAGGTGTGTACCACA
CCGATTTTATCTCTACGTGGATGAACCACAGCTGTCGGCTCCCTTGTCTC
CAGGACATCACACTCTCCACATTCCCTCCCATCTTCCGGCTTCTGCTTCC
CGGGGCCCTCATCTGCCCCATCCTGGGTGAACACTGGTCGGTCAACTGCT
GGGCGTACCTTCCCGCTCTGCACACCCTCCCTGGCCACCCCACCCACTCT
CACGGCTCGCACTGCAGAGGAGCCGCATCTCTAGCTCCAGCCCATCTGCC
TCTTCTGAGCTCTAACTTCATGTAGGCGACTCCTGCCGGTGTTGCCTCAC
AGGCCCATCATACTTCAAAGCATTTTCCCCTCAGAACACCATGTCCTGGC
TGCTCCCTCCAGAAGATACATCTCTCAAGCACATCCCCGCGGCTCTCACC
TGGATGACTGCATTCACCTTCTCCCACATTTGCCCCCTCCTTTGGATGTA
TATAGATTGTTTTAAAATACAAATCTGATGTGCTTGCTCTCCTGCTTGAA
ACACCTCAAAACTGCCTTCAGGATAAACCACTGCCCTTGACATGTTCACA
GGTTGCCCATGGCCTGGCCCTGCCCATCTCTTCAGCCTCATCTCATGCCC
CTTGCCCCTCGCTCTCTGGGCTTCTGCCTCCCTAGCCCTCCTTTAGGTTC
TCTAACACACCATAGTCCTTCTAGTGTTGGGGCCTCTGCAAGTGCTGTTC
CCATTGCCTGAGACATGAATCCCTCTCCCTATCTCTACCTGCACCTTCAT
CTGATTAATCCCTACCCTTCCTACTCATGATGTTGCTTTCTCAGGGACTC
TCTCTGACTTTTTAAACTAATCAGGGTCTCCCCAGTATATATCTTCATAG
CACTCTGTATTACTCCTTTCTTAATGACCACCTGCTGTAGACAGAATGTT
TGTCTTCCTCCAAAATCATATGTAAAACCTTCCACCAGAGCGATGATTAG
AGAAGCCTCCC

>Contig37

GACTGACATTCAGAAGATATTAATAAGAGCACTAATGATGGGGATTGCAA
CCATGTCTTTACTGACTTCCAGAAGCTTCTTACAGTAAACATGAAATCAC
ATAATTTCTTCCACTTTCCTACTGTTTCTTGTTCTGGGCTCTGTCCTGCT
TACTGTCTAATATCTTGGCCCCTTAAAAGTTGCTAATCTTCCAAACCTCA

FIG. 4 (14 of 61)

TTCCTGTGACTGGGCCG.TGGTCCTTG..CATGGGCCTTGAAGATACTGA
CTGTACACTTATCTGGAGCATCCAGTGCCTACCACCTGACCCAGATTCCT
CATTGCGCTCCTCCCTCCTCCACCTAATGGGATTTGCTCATACCCGTGTG
GGACCCCTCCCATTTTCCCCAACTGAATACTTATCAAGACAACGCATTGC
CATACTCCCTCGTACCCTGCTCTGGGCATCAGACTGAATGTTTGTTTCCA
TTGAGGATCTGCAGCTGCATCAGTTTCCCCAGCACCGTCCAACCCCTTGA
GCATGGCTAGTCCTAAAGCAGAGAATTAGCCTTTCTATCCCTGCTGCTAT
ACATGCTGGGACAAATAATAAGAAATGACAGCATTTTATGATAATGCAGG
CTGCAGGAGGCAGGAGGCAGGAATCAAATTCGTGCTTATCAAATAGTGCT
CCAATTCTTTGAATATTGGACTATAGAATATGTCATGGATCTATGCTCAG
GTGGGTTCCCTATTACTCACTCCACTGAGGCCAGGTTGTGGGATTAGCTG
TCCAAGAGGGAGTTTCAGTCTCACAGCATAGGGTCATTCTGAGAATTACT
GGCCCACACTTGTGTGGAGACCTCCAGAGAACAGAATCTGGGTTGGTGCC
ATGTACTTCCAGGAGGAGAGAAGTGGCAGGATGCCCAGCCCCACAATCAG
AGGGGAAGGGGCAGAGCCACATGTATGAAGATCCTCTCCCCAGTACGTGC
CAATCACAGGGCTTCCTAGCTTTTGGGCCAAGGAAACAATGTGGGAAGCA
AAAAAGGACAATTTTCTCCTCCCTTTGCATGAAGACTGAGCAGTTTTACC
AGATTCCCAGGGAAACACCCTTCCACTCTGGGTTGAATGTGAGTGAGAGA
CATTCAGCTGGAACACTAGAAAAACTATTTCCTGAGCCACTCACCTTTAG
CCCTAGAAAGTGTTGGATTTGTCCTTCATCTTTGCCACAGTAGAGACTGC
TGATAGCATCAGAACTTGGGCTCTGGAATTAGACAGATATGGGTACAAAT
CTGAGCTCTCTCACTTATTAGTGTGGGATGTAGAGCAACTTTTAAAATCC
TTCCAAACCTCAGACTTCTCATGCATGATGTGAGGATTGTAATAGGGCCC
ACCTAATAGGGGTTTTTGAGAATTAAAAAAGTTATTCAATGAACAGCATT
TAGCAAGATGCCTGACCATTGAGAAAATAACAAATTGTTTATTATTATTG
TTATTATTAAACATCTTTCCTGCACCTTCTGACTGGGGGCATCGTATCAT
CAGAAATACTTAGGATGGGATGGATTCCTGCATGGGCTGAGTCAAGGGTG
CAATAATGGAGGAGTGAAGAAGGAAGAAATGGAGGCAGAAATCCCCAGGA
GCCCAGCATGGTACAAGGCTGAGCTAGTGCTGCAGAGCCTCCTTGGAACA
GCCACAGAGCTTGCATCTGGCCCTGGAGGAACCTCTTCTAGCTGGCAGGA
CCAGCCACAACAGTGGCCAGGGGATTTCCCAGGGCGTGGGCTCCTCAGGA
GTTCATTTGGACCAAGCCTGCCTGGAGAGGGGTTATAACAGGGATCCTTC
CCTACTGGCAGGTGATTTACCCCTCGGTGAGAAGCTCAGGCATTTGTTTG
ATGGAAGGTGGAAGGCCCTGTGCTGGGCCAGTGACTATCAGGGATGGGCG
GGTGGCTGGAAAATAGCAAATAAGACAATATGATAACACAGTTAACCACC
ACACTATGTGAAGCTACAATATGGGTATCTGTAATAGACAATTCCAATGT
AGAGAATAATTTTAAGGTGTCATTCTCCCCGCCAATGCCATAAGCACACG
GCCTCTGCCTGGGTTTCTCACTGTGGAATGTCCTCCTGGTCTCCTCATGC
CCAGAGAGTGGGAAGTACTCCTACTTTAACACCGGCTTTCCTGTCATTTC
CNTGCAGCCCTCCTCAGCCCCCTCTGCACAGGGAGGTTTCCTCCCTGCTG
CTGCAGTGCTTTGTACTTGTTAGTGGTACCTGCACACAGGTATTGGTGTC
CTTGTCTCACCACCCTACATCACTGTAAGCTCCCCAGGAGCAGGCTTCCT
GTTTGACTCACCTGTGATCCTCCACCTCCCACCCTGTAGTGCCTCAAGCA
TTCTGTAGAGCACATGGACGCC
>Contig38
GACTAATAAGTACTTCATTATTTGGGTATTTTCCAAGAACAACATATTGT
AGGAAACCATTCTTTCTAAAAAAAAAAGTGTCCTTTTAAAAAGGTGAATA
ATTTTTGTCTAATTCAAAGTTTATTGAAAAGTTATGTATAAAACAAGGTA
AAAGGAACAAGGAAATAAGGGAAATGTAAAGAAAATTATAGAAATAAAGT
GGTATTTTTTGGTAAGAAAGCTTAAAGAGAAATAATTTTAGGTAAGAAAG
AATCTTACCTAAAATTTTGTGCTAGAATAAAGTGACTGGCTAAGAAAGGG
ATGTTCAAAGCTATTTATGACAAACCCACAGCCAATATCATACTGAATGG
GCAAAAGCTGGAAACATTCCCTTTGAGAACTGGCACAAGACAAGGATGTC
CTCTCTCACCACTCCTATTCAACATAGTATCGGAAGTTCTGGCCAGGGCA
ATCAAGCAAGAGAAAGAAATAAAGGGTATTCAAATAGGAAGAGAGGAAGT
CAAATTTTCTCCGTTTGCAGATGCATGATTGCATATTTAGAAAACCCCAT
CATTTCAGCCCCAAAACTCCTTAAGCTGATAAGCAACTTCAGCAAAGTCT
CAGGATACAAAATCAATGTGCAAAAATCACAGGCATTCCTATACACCAAT
AATAGACTAACAGAGAGCCAAATCATGAGTGAACTCCCATTCACAATTGC
TACAAAGAGAATAAAATACCTGGGAATACAACTTACAATGGACATGAAAG

FIG. 4 (15 of 61)

ACCTTTTCAGGGTGAA..GCAAACCAC..CTCAAGGAAATAAGAGAGGAA
ACAAACAAATGGAAAAACATTCCATGCTTATGGATAGGAAGAATCAATAT
CGTGAAAATGGCCATACTGCCCAAGTAATTTATAGATTCAATGCTATCCC
CATCAAGCTACCATTGACTTTCTTCACAGAATTAGAAAAAACTAATAGCC
AAGACAATCCTAAGCAAAAAGAACAAAGCTGGAGGCATTGTGCTACCTGA
CTTCAAACTATACTACAAGGCTGCAGTAACCAAAACAGCATGGTACTGGT
ACCAAAACAGATATATAGACCAAAAGAACAGAACAGAGGCCTCAGATATA
ACACCACACATCTACAACCATCTGATCTTTGACAAACCTAACAAAAATAA
GCAATGGGGAAAATAATTCCCTATTTAATAAATGATGTTGGGAAAACTGG
TTAGCCATATGCTGAAAACTGAAACTGGACCCCTTCCTTACAACTTATAC
AAAAATCAACTCAAGATGGATTAAAGATTTAAACATGGCTGGGCATGGTG
GCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGATGGGTGGATCAT
GAGGTCAGGAGATGGAGACCATCCTGACTAACACAGTGAAACCCTGTCTC
TACTAAAAAATACAAAAAATTAGCTGGGCATGGTGGTGGGCGCCTGTAAT
CCCAGCTACTTGGGAAGCTAAGGCAGGAGAATGGTGTGAACCCAGGAAGT
GGAGGTTGCAGTGAGCCAAGATCACGCCACTGCACTCTAGCCTGGGCAAC
AGAGTGAGACTCCATCTCAATAAATAAATAAATATGGAACTCTCCCAACA
CAATAATAAGACAAACCCCCAAATGTTTTAAATGGGCAAAAATATTTGAA
CAGACACTTCACAAAAGAGGATATGTAAATGGTCAAAAAGCACATGAAAA
GATGTTCAACACCATTGGTCATCAGGGCAAAGAAAACTAGAACCACAATG
AGATGCCTCTGTACACCACTTAAATGTCCAAATTAAAGAAAACAAGTTTT
GGCAAAGTTGTGGAGCAACTGAAATGCTCGTGTATTGCTGGTAGAAAAAC
AAAATGGCATAACCATCGCAGATAATTTGTTGTCAGTTTCTTACAAAGTT
AAACATATACTTATTGATATGACAGTTCCATTCCAAGAGAAATGAAAACA
TAAGTCCACACAAAGACTTGTACCTGGGTGTTCATGGTAGCTCTATTCAT
AATTGCCAAAATCTGGAAACAAATCAAATGTCCATCAGCAATGGAATGGA
TATACAAATTGTGGTACACATGTACAATAGAAAACTACTCTGCAATGGAG
AGAAATTAACCATTGACAAACACAAAAACATGGACAAACCTCAAAAACAT
TATGCTGAGCAAAAGAAGCCAGACACAAAAGACTGCTCAGCGCATGATTC
CATTCATATGAAATCACAGAAAGGGTCAGTTGAAGGTGCAGAGACAAAAA
GTAGATCTGCAGTTGCCTGGGGATGGGGTGGAGGTTGACTGCTCTGACG
CGTAAGGAAATTTGGGGGTAGGTGGGGGATGGTGGGAATATTTTTTGAAT
TGAATTGGGTAATAGTTTTAATAGGTAAAATATTGGACCCCACAGTATTT
GAGATAGGTTTCAGTCAATTTAGACAGTTTATTTTGCCAAGGTTAAGGAT
GCATCCGTGACCCAGCCTCAGGAGGTCCTGACAACCTGTGCTGAAGGCAG
TCAACATACAGCTTGCTTTTATTCATCTTAGGGAGACATAATACATCAAT
CAATGCATGTAAGGTTTACATTGGTTCAATCTGGAAAGGTGAGGGAACTT
GAAGCAGGGAGCTTCCAGGTTACAAGGTAGATTATTCTCAACAGAAAGGA
ATGTCTGGGTTATGATAAGCGGTTGTGGAGACCAAGGTTTTATCTTGTAG
ATGAAGCCTCCGGGTAGCAAGCTTCAGAGGGAATAGATTGTCAAAGTTTC
CTATCAGACATAAGGTCTGTGTTGATGTTAATGCTGGTCAGCTTTTCCTG
AATTCCAAAAGGGAGAAGGGTATACTGGGGCATGTCCAACCTTCCCTTCC
ATCATGACCTGAACTAGTTTTTTCAGGTTAACTTTGGAATGCTCTTGGCC
AAGAAGAGGGGTCCATTCAGATGGTTGGGGGGGCTTAGAATTTTATTTTT
GGTTTACAGTGAAGACTTTTCAAGCTAGACACTTAAATGAGTATGTTGCA
AAATGGCAATTTCTTAGCACGGC

>Contig39
GACGTCCTAAAGAAATGCTAAGGTAACTCAATTAACTATGCTAGAAAAGA
GAGTTAAGTATTTAGGAGGATTTAATATGGTGTTAAAGTTGTGAAAATCA
AAATGGAGACACTAATGTTAAGAAAACCCTGATAAATGGAGCCAGGGAAG
GCCATGAAGAAAGAGTTCTCACACTTGTATCCCTGATCATGAAAAAGACT
CTGCAAAAAACAAAACCTTGCACAAAGGCCATTGCAACCTTACACAAAAA
ATACTACTTTAAAAGGACATGTGCCCAGCAACTGCCTGTCCAACCTCAGA
CTGGCAATATCTTTGTTATTGATCTTAGTAGCCCAGCATAACTATTTCAA
AACAGTGATGTAATGCTCATTTTTTTTCTTTTGAAAACTTTTGTCTTCCT
GTAAAAACCTTTGTCTTCTTTACTTACCCTGAATATGCACAGAGTTTACT
ATGGAGTGCATATTCCTGTTGCAATGCTCTATTCCCAAACAAACATCATT
TTCTTTTAGAGAGCCTCTCTCTGTTTGTGATTTAGGTTGGTGATGTAAAG
CAATGGCATAACTGAACACTGATTCAAAGAAAAGTGGCTTTTCTCTTTGT
TGTATTAAAAAGAGGCCTTATAAATAGGATAGTAAGATTTGTAAGTTGAA

CTTAAAGCATGAAGAAAATTTAGGGGCCAGGCAGGGTGGCTCACACCTGT
AATCCCAGCACTTTGGGAGGCCAAGACAGGAGGATTGCTTGAGCCCAGGA
GTTCAAGACCAGTCTGGTCAACACAGACCTCATCTTTACTAAAAATAAAA
AAATTAGGCCAGGTGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA
GGCCAAGGCGGGAGGATCACTTGAGGTCAGGAGTTCGTGACCAGCCTGGT
CAACACGATGAAACCCCATCTCTACTAAAAATACAAAAAAATTAGCTGGG
TGTGGTGGCGGGCACCTGCAATCCCAGCTACTCGGGAGGCTTCAGGCAGG
GGAATCACTTGAACCTGGGAGGCGGACATTGCAGTGAGCTGAGATAGTCC
CACTGCACTCCAGCCTGGGCGACTCAGCAAGACTCTGCCTCAAAAAAAAA
AAAAAAAATTAGTCAGGTGTGGTAGCACACAGCTGTGGTCCCAGCTACTC
GGGAGGCTGAGGTGGGAGGATCATCTGAGCCCAGGAGGTCAAGGCTGCGG
TAAGAGCTGAGATTGTACTACTGCATTCCAGCAGGGGCTACAAAGTGAGA
CCCTGTCTCAAAAAAAAGAAAAAGAAAAAGAAAATTATGTTTTTAAATTTA
TAATTATAATAAATTTAATTACATAAATTTAAGCTCAAGTAATTGTAAAT
ATTCTTTCTGTGCACATAAGTTATTCTTGTATTGACCCCACAGGAGCTGG
CCATTCTTCAAGTCAGAAGGCCTGAGAGAGGAGCTGCCCAGGTGGTCTTC
ATGGGGCTGTGCGGCCAGTCATCCCCCACAGGTTGACAATCCTTGTGTAC
TTCATCCTCGTTGGATCCTCTGTATCCCTGACGATGAGCAACTGTGAGGC
CCGTTTCAGCACTGAGTTCCAGTCAGGAAAACATCCACCCACCCACCACA
CGCTCACACTTACACACACATTCACACATGCACACACGTTCTGGCTCCGA
AAAAGAAAAAAAAAAAGCAATTTAAAATAATTCTGATCCTTTGCTTATTT
CCACAAACTCCATGAAAATTGTACATTGTCCAAGCAACATTTCTTAATAT
TCTCTTTTCTCTCATATCCATTTTCCTTACTGCTGTCTCCACCTTTCTC
TTCCAAACTCCCTGTTAAAATCCCTGCCCCAGCGAACTTTTATTCAATTT
TGTGGAATGGAGGCTGCTCTGATTTAAATTAAAAAAAAAAAAAAAAATCCC
TACTCCATGTCCCAGATCCCTAGTTGTTTTTTGTTTTTTGTTTTCCTGAG
ACAGGGTCTTGTGTCTTCCATGCTGGAGTGCAGTGGCATGATCATGGCTC
ACTGCAGCCTCAACCTCCTGGGCTCAAGTAAATCTCTTGCGTCAGCCCTC
CCCAGTAGCTGGGAGTTCAGGTATGTGCTACCATGCCTAGCTAATTTTTT
TCTTTTATTTTGTAGAGACACGGTCTTGCCAGGTTGCCCAGGCTGGTATA
GAACCCCTGGGCTTAAGTGATCCTCCTGCCTCGGCTTCCCAAAGTGCTGG
GATTACAAGTGTGAGGCACTGCACCCAGGCTGGATCCCTGCATTTTTACA
GATTTAGCATCACAAAAGTCTAAACAATTAGACTGACTAAGGCAGAACTG
CCCTTATGACAGCAGACATAAGAAGGAAAAGGCCAAAACACTGTGTTAAA
AATTATCCAAATGTGAGGAAAAGGCAAAGAGAGTAGGTGTGCCTTTTTAG
TGTCTAAGCTGCCTGCCCAAGGGGCATCTGATGCTCTCAGGCAGGAGTCC
ACAAATTTTTTTTTGTAAAAGATCAGATAGTAAATCTTTTCAGCGTGAAG
AGCATGAGGTCTCTGTCACAAATACTCAACCACCATTACAACATGAAAGC
AGCCAACAGACAACACATGACAAATGAGTGTGGCTGTGTTCCAGTAAATC
TTGATTACAAAAACAGGCAAGAGGCCAGAGCTGACCCATGGGCCATAGTT
TGCTGACCCCTTCTGTAAAGGAAAGTATTTTTGTTTGACTTGCTGTTTAC
CATTGATTGAACACAAGGCTCTGTAGAGTTACTTGTTAACTTGCAGAAGA
TTGATGAGTGGCAAGTAATTTTTATTCACCAGAATATANNATTATTCTGT
TCAGTAGATAAGATAAACCCACTGTTATATTACTGTCTTGTTTAGAATGT
GACTTTGATTCATTTTTTCACAAATTCATATTATTGCCCTAATTTGTATA
TAAGTATGCTTCTTTTAAAAATATATATTTTTTAATAAATTTGAGACAGG
GTCTCACTAGGTTGCCCAGCCTTTTGCTATAATGAGAGCATAAAGTGAAT
TTCACACTTTAGCCTAGTGCATAGATGGGATTACAGGCACAAACCACTGC
ATGCAGCTAACTTTGCTTCTCATTCCAGCACGTTCTATTCCNNNGNTTTT
CATATACGCGTCTCTTAATGC
>Contig40
CGCATTCAGCCCAAGTTTTCTTCAGTGTTAAGGTTTTTGTTACTCTGTGC
CCAAATGTCCTTCCAAAAAGGTTAAGTTTTTTACCTTCCTGCCAACATT
ATATGAAAGTGTCCACTTTTGTAGACTTTTACCAATGCTGACTACTTTTG
GTTTCAAAAAAGCTCTCAGTAATTTTCTATTAATTACTTTTACCCTTTTT
TATTGAGGGTGTTCAACTTTTTATTGTTAGCATATTCTCTCTGGGCTCCA
TTGGACGCCTTGGCAGCTTTTTGGTAGTAGGTGCCTTTAGAAAAGTCCTT
CTCGTCTGGCCCTTTCTGAGCAAATCTAGTGAACAGAATTGGCTCCATGC
TCAGCATTGCTTAATACGGTTGATCCAGGGCCTAGGACTCATTCCTTCAT
TACCATCCACTTGCATTGTCTTAAAGCAAGGCTCTATTAATTTAATTTGG

CATTTCCTGTCCCAGCTCTTTTAGTTTCATTAAACAAAGGCTTTAGAAAAC
TCCCAGTAGATGCCTATGTTGCTTCCTTTTAAAAAATTTTGGAGCTGTTT
CCCTAGCCTAACCTTTTCTTCAGGGCAGGAGTTAAGTCCCTTCTACTGCA
TTCCTGTGAAGATGGTGATTCAAGAGGCAGGGCACCTGTTGCTTTGTGAA
ACAGTCCACTCTGCAGCTGGGCAGCTCTGTTACTAGAATGTTCTCCCTTC
TGGGGAGCCAATATTTTGATGTCCTCTGTGAATCTCATCTGCTTATCCCA
TCTGTTTATGTCCTTGAAGATGCACAGGTCTGACACCACGAGGTAGCCCT
TAGAAATTTGATGGCATTTCTGATGTGTCCCCAACTCTTCTCCAACCACT
CCTCCCAGAGCTTGTTTCTTAAGCCCCTTGTGGAGCTGATTGCTTTCCTC
AAGGCAGCTCAGTTTTTCCCAGTTTGCTCCTGGTGGTCCTGAAATATGAT
TGACTCCTGAATACTCCAGGTGTGAAGGAGAGTGGGGGTGGCCTTTCTAC
TTGTCATGGCCTGGGTTTTAAGTTGCTGTCCAGTGGAGCAGAGGTGACTT
TCCCAGTGAACTACATTTTTTCCCCTCTAAATCCTTAGCAATTTTGTCTC
CAGAGGCAAGACCTGGCCAAACCATTTGTGTTGAGGATTGAATCAAGAAT
GATTGAGGAGATGACAGTAGTCCCCCCTCATCTGAGGAGGGCGTGTTCCA
AGCCCCTCAGTGAATGCCTGAAACTGTGGATAGTACCCAACTCTATATGT
CTATGATTTTCCTATAAATTAATACATGCCTGTGACAATGTTTAATTTAT
AAATTAGGCAAAGAGGCCAGGCGCAGTGGCTCAAGCCTGTAATCCCAGCA
CTTTAGGAGGCTGAGGCCTCACCTGAGGTCAGGAGTTCGAGACCAGCCTG
ACCAACATGGAGAAACCCCGCCTCTACTAAAAATACAAAATTAGCTGGGC
ATGGTGGCAGGCGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG
AATCACTTGAACCCGGGAGGCGGGATTTGCGGTGAGCTGAGATCGTCTCA
TTGCACTACAGCCTGGGCAACAAGAGTGAAACTCCGACTCAAAAAAAAAA
AAAAAAATTAGGCAAAGAAAGAAATTAACAACAATAAGTAATGAAATAGA
ACAATTCTAACAATATACTATAATAAAAGTTGTATGAATGTGGTCTCTTT
CTCAAAATTACCTTTTTTTTTGAGACAGGGTCTCACTTTATTGCCCAGG
CTGGAGTGCAGTGGCACGATCACAGCTTACTGCTGCCTCGACCTCCTGGG
ACCAAGTGATCCTCCCACTTTAGCCTCCTGAGTAGCTGGGACCACAGGCA
TGCACCACTGTATCTGGATAATTTTGTTTATTTTTTTTTGCAGAGAGAGG
AGGTCTCACTATGTTTCCCAGGCTGGTTTTGAATGCCTGGGCCCAAGGGA
TCCTCCTGCCTTGGCCTCCCAAAGTATTGGGATTACAAGCGTGAGCCACC
ATGCCTGCCCCAAAATTATCTTATTGTTCTATACCCACTCTTCTTCTTGT
GATGATGTGAGGTGATCCATTGCCTCCTTGATGAGATGAAGTGAGGTGAC
TGATGTGGGCATAGTGATGCAGTGTTTAGGCTGATATTGGCCTGATGATA
TGTCAGAAGGAGGGTCATCTGCTTCGGTGATCCTGGATCATAGAGTCATG
ATGATGTCAATGGTTGGATGTCAGGAGCAGACGATGTCAATGACTAACGA
TAAGCTGGACAGGTGGGATGGTGGCACAAGATTTTATCACGCTACTCAGA
ATGGAGCACAATTTAAAACTTCTGAATTGTTTATTTTTGGAATTTTTCAT
TAATATTTTTGGATTGCAGTTGACTGTGGGTAACTGAAACTGTGGAATGT
GAGACTGTGGAAAAGTGAGGGAGTACTGTATTATGGAACTGTAACTCTAT
TCGGTAGGGGAACAGAATTCACATTTGTGGGGCCCAGGTCTCTGCATCTG
TAGGGATCCAATTGTTTCATTTCTCGTTGTAGCAAAAACTTGGCTTTGGA
ATCAGACAGATTGATGTTTGCTATCATTCTAAATGGGTGCAGCTACACTT
TCCTCAAGAGGTAGTTCTGAAAATTTAACAAAATGTGAATTTCTTGGTAA
AAAAAAAAAACCTCAAAAATATTCAGTTTCCTTTCCTTTGTGTCTGATGT
ACTCCATCAAATACTGGGAAATATGTGTCTCTCATAGAAATGTCATGGAT
CTTTGTAATTCTGATTATCCACAAACCTTGGGGATTAGCTGTTTCAATGT
TCCTATTTTACAGATAAGAAAATGGAGCCTGTGGTAAGTTAAGTGAGTTA
CTCATGGCTACTTAACTAATATTTTACTAGGTGATAGGCCAGAGCTAGAG
CCCAGGTCACCTTCTTATCAATGCTCTGCCTTGTCTCTGTGCCTTCCTGT
CTGTCTGTATGTGTATGTGCCTGTTGACAGTAAGGCATAGTTTAACCCAG
TAGAACTACCGGTTTGTAATGAATTCCACTTGTAAATGACTGACCATTCA
AGGAACAAGTGTTTTTTCTATGCTTGACACCTGTTTTGGATGCCAAAAAG
GATACAAATGTAACTTCAGACACTCTGGGCCTCATTTTGCACTCATTAGC
ATGTCCAAAATTAAAAAGACTGACCACACCAAATATTGGTGAGGATGTGG
AAGAACGGGAACTTTCATACACTGCTGGTGGGGATGTAAAATGGTACAAT
CCCTTTGGGTAACAGTTTGACAGTTTCTTAAAAAGTTAGACATATATATT
TACCATATGACTCAGCCCTTCCACTTCTAGGTCTTTACCCAAGAGAAATG
AAATGCTGTGCTTTTACAAATGTCTATACAGGAATGTACATAGCAACCTT
ATTTGTCATTGCAAAAAACAGAGACAATTCAACGTTGTCAAGAGTGAATG

FIG. 4 (18 of 61)

72/118

GATGAGCAAGCTGTGGT. GTCTATGCA..GGTATCCTACTCAGCCAG
AAAGATATGGCTAAT
>Contig41
GACAACAATGTCATGCATAAGATGACGATGGCCTGGGTGATTGATGCAAA
CAAGGATAAAGAAAATAATCAATTTTGTCCCCATTTTCAAAGACAGATAG
CAGCAGCAAGAGTGTAAGTCTGAGGAAAGTCATATTCCTTCCTCCTACAA
CATAGCACACACACTTACAAAAACAATACACAGACTCCTGGCCAATGGAC
TTCAAAACTGAGGAGGATCATTAAATTTAAATGTTCACCGCTGCATGAAA
TCTCCCTGGGTCCTGCCCTCCCTTCCCCACCCTCCTCCACTTGGGCCGGG
GCACAGCAGTGATTCTCTCACCTCTCAGAGTGAGCCAGTGTTGGCTGCAT
TGAAGGCTCCAGATATGCAAACAGGGCAGATATTCCTGGACCAGGGTGCA
CAGAGTGAGGCTCCAACGCACCCTATTAACTGCATGAAGGATGAATGAGC
CTCTGGTATGGGCTGGGACAGAAAAAAGGATTCAAGGGGCCCAAAAGGGT
TTGGGTGGAACCTACCAGGAGCGGCAGTACAGACTCCTTGGGAAGGTGGC
CATGATTTAGCCACATTCACCAATAGGATAATCTGGAGAATTTCCTAGCT
TGAGTTTCTGGGAGAAAGCAGATTTCTGGATTATCTGGTGACAGGTAACA
GGGCCGAGTTCATCCACAGCCACCTGCAGTGTTAGCACCTTAAGCTGAGT
TCCTTGCACCAGGATGCTGTCACGCCCAGTCAGTGTGAGACGGTTCTTGG
CTGAAGGACTGAAAAGCTTGGGTAAGTGACTTCACCTAAGCCTCTATCTC
TTGCTCCCGTAAGTCAGGGCTCATTGTGGCTCCTTGCAGGCTTGACTTCA
GGGTTAACAGAGAAAATGAAGGTACAAGTGCCTTGTGAACTCTGAAACTC
CAAACCAGTCATTCTCAAAGTGCCGTCCACCAGTCTAGCACATCAGCATC
ACTGGAAGCTTGTTTGAAATGTAAATTATCAGGTCCTCCAGAGCTATGTA
TGAATTAGAAACTCTGGGAATGGGGCCCTGCAATCTATTTCAACAGGTCC
TCCAGGTGATTCTGATGCAAGTTAAAGCCTGAGAAACTCTGTCCTATACA
AATGGATGTCAACTCAAGCTGCTCTTCAGAATCACCTATAGCACTTGTTC
ACCCGAATCCCTGAGAATGGAGCTTCAGGACTGCTATTTCTCAAAGTTTG
CCTGGTGATCCTGAGATGGGGTTTGGGGGACAGAGATCCAAGGTGCTACC
AGGTGTGAGGAATTGTTAGAAGGCAAACCTGGCTGTCATCTAGGGTGCTT
AAAGGGTACAGATCCTAGGATTCTGCCTCTTACAGCTGAATCAGACTTTC
CTAGAATGGGATTGCTGTCCAATGGCATGCCTCCTGGGTGACTCTGATGT
ATAGCCTGGGCTGGGAACCACCAGAGGATTATCTTCCATTGACCAAGCTG
ACAAACTCGCTTAAGGCTCTGAGTTTCACACTTGATTTTCTAGCCCCTGT
CCTTCCATGGATCACCTGCCCCCTTCCCTCCTAATCAGGAGCACAGTCAG
TGGATGCACTAATGTGGCCTCTCCTTGGCTGCAGGGAACAGGTGGAAATG
TGGCCATAGGTGTGCAGGGCTGCCTGCCATGTATTAATAGCTACAGATTT
GAAAGATCCAAGGACAAGAGACTAGAAAAAAATTTAAAACAGCCAAGCAT
TGGCCCAGTAATGGCATTTCAGAAATCCACCAAAATATTAAGATGCTTTT
TGAAAAATATCCAGAGCACTCATGTAAAAGTGCTTAATTATTAATAAAAG
CTGACATGTGTTGGGTACTTCCTGTGGGTCTGGCACTAGGCTAATTATGT
TTTTAGGAGTTGACTCAAATGCTCCCTGTCATAATTATGTGAAAAAATAT
AATTATTAGCTCCATGGTACAAATTAAGGAGAGGTTACATAAATAAAAAG
GAATGATACTCAAATTAGTAACCAGAGCCCATGCTCTTAAACACTATGCT
ATTATTTGTGGACTCTTACATAGGTGGCAAAAGTCAAAGGCTAGATTGAC
TTCTGTCCACTTCCAGCCAAGATGAAGTACAAGATTCAGATACACCCTTC
CGCATTAAACAACTTAGGAATCAGACAAAATATACAAAGCATTGTTTGTT
ACACATTGGATAACAGACAGCACTAGATAGTCGTGTCTGAGAAAAGCGGT
GAAATGAGCTGAGTCTTAGAATTGCCCCAGTTTACTAAGGGGCATAGTAA
GGGCATAGCTGCAGCACAAAGAAGCAGAACCCAACAGAGACTGGCGTTCA
CCTGAGTTGAGAAAACCAAGTTGAAAATTTAGGAACACTAACACAGATAT
GTAGGCAAGAGTATCAGAGAGGAGACAGTTGTAGGGAAAAAGAGAGCTTT
ACAGAGAGACAGCGAGAGCTCCAGAGACCCGCAGAAGATTGCCCTGACGT
CACTAGCTGAGTACCGATCAGTGCATACATGTAAGGATATTACTCAATAT
GTGGAAAAGAACAGAAGGAATGATGTCCAAAGCTCACCCAAAGACAGGAA
TCATTTATGTTTCCACCAGCCAGAGTGGAACAACCTTGTAACGCATATGG
AGTACTCAAACGAATATTTCCTCAATAATAAGTTCAAATTAACTGAGACT
AAAGCCTGCCCGCTTTGTCTGGACATGCCTAACAAAGCTTTGAGGGAAGC
CTCAAAAGAATGAAACCGTGTCCAAGTAATTTAACTGTGTCCCAGAAAAA
AATTCAAGAACATTTAAATAAATATTAAAATATGATCAAACCCAGCAAGG
TTAAATTCAAAATGTCTGGCATCCATTAAAAAATTACCAGCCTTGAAAAT

TGGCGGGAAAATATTA. .ATAATGAA. .GAAAAAGCAATCAACAGA
AGGCCTAGAAAGTATACATATGATAAAATTAGCAGACATTAAATGGTTAT
GATTAATTTATTTTATATGTTAAAGAAGGTAGAGAAGAGCATAAGCACAT
TAAAGAGAGACAGGAAAGTCCCAGTACTCACACAGGGCCAGGAGCAGTTT
TCACCAGTCAGGTGGGAAAACTTCATATTTCATGGAGCATTGGTAGAGTA
CACAGTGTCTTGCCTTAGTAGAGGGATAAATGCTGTTCTGTTCCCGCCTA
ACCCATCTTGAAAGAAAATCTGAAAGGATCAAACTGTATTCAAGTAACCT
AATCACATCCCAGCACACAGCTCGACTAGTTATAAAAACACAAAATATTA
ATATCTAGAAACACAAAAATAATATCTAGCACCCAACAAGGTAAAATTCA
CAATGTCTAGCATTCAATTGAAATTTTCTAGGCCATCAAAGAAGCAGTAA
AATATGACCTATAAGGCCGGGCACATTGGCTCATGCCTGTAATCCCAGCA
CTCTGGGAGGCCAAGGTGGGTGGCTCACCCGGAGGTCAGGAGTTCAAGAC
CAGCCTGGTCAACATGGTGAGACCTCATCTCTACTAAAAATATAAAAATT
AGCCCAGCATGGTGGTGGGCGCCTGTAATCCCAGCTACTCAGGAGGTTGA
GGCAGGAGAATCGCTTGAACCTGGGAGAAGGAGACCGCAGTGAGCCAAGA
TGGCACCAATGCACTGCAGCCTCATTAGAGAACATCGGGAAG

>Contig42
GAAACTAAAGGCTTATTTAAAGCGCGAGACCGTGGCGCCTTTGGACTGGA
CCCTTTCTAATGATCATTTAGTATCAGGCTATGTGGGAGTTGACCGTTTT
GCATAGCCTGAAAGCCAACAGTATCACTCCTCCTCTAGGTGTGGCAGAGA
TGTGAGAGAAGGAGACTGACAGTCTGTGGGTGTGTATGCAGTGTTGGGGG
AAGCGAGGCACAGGGGACAATACTGTGGTGTAGAAAACTAGTCTAAGGTA
GCATCAGGAAATTCATGAAACCAAAATGAATTTCATAACAGCACAAGACA
TTATTTGTTTTTGCCTCCCTCTCATTTTTTTTTTTTTTTGAAACAGAGTC
TTGCTCTGTCATCCATGCTCGTGTGCAGTGGTGCAATCTCGGCTCACTGC
AACCTCCACCTCCAGGGTTCAAGCAATTCTCATGCCTCAGCCTCCTGAGT
AGCTGATTACAGGTCTGCACCACCCCGCCGGCTAGTTTTTGTATTTTTAG
TAGAGATGGGGTTTTGTAATGTTGGCCAGGCTGCCCTGTCATTTTTTTTT
TACTAGTGTCCAGTGGAGTTTTTAGGGGCTACATAACATGATACTGTCA
TTAATCTAATGGCTAATGAAAGGGATATGTATATGTTTTTGTGTTTAAAA
CAAACTTCTTTGGGGTCCTCAATAATTTTTAAGAGTATAAAGGGGTCCTG
AGATCAAAGAGTTTGAGTTCTGCTGGACTGGGACAGTGGTTGTCAACCCA
GATTGTACATTAGGGTCATCTGGGAAGCTTTAAAATAGTACTGATGCCCA
ACCTTACCGCAAACCAATTAAGCCAGAATCTCTGTGGATGAGAAGTCTTC
ATTGTCATCATCACCATGACCATCATCATTGTCACCGTCACTACACCATT
ATCATCATCATCATATCATCTTCATTATCATTGTTAGTATCTCCATCACC
ATCATCAGCATCACCATTATTATCATCATCATCATCCCCACCATCATCCT
CATCGGAACTTCACCTGCATGGAGGACAATCCACTATGCATTAGGTGCTA
TGCTATTTGCTATACTCCTTATTCTCACAACTGCCCAGAGAGGCTGATAT
TATCTCACTTTATAACAGGAGGAATCTGGATCGGAAAAGTTAAGGTAAGC
TAATTCACAGAGCGAGAAGAGATAGAGCCAGGATTCGAAACCAGTTCTCT
GCTACATCAATGTTCCCAGTCCTTGCACTATTGAGAACCTCTTTAGTTAT
GCTTTCACCCCTCCAACACCACAGTAAATTTTTTCTTTTTTTAAAAAAAT
TATACTTTAAGTTATAGGGTATATGTGCATAATGTGCAGGTTTGTTACAT
ATGTATACATGTGCCATGTTGGTGTGCTGCACTCATTAACTCGTCATTTA
CATTAGGTATATCTTCTAATGCTATCCCTCCCCCGCTCTCCCCACCCCATG
ACAGGCCCTGGTGTGTGATGTTCCCCACCCTGTGTCCAAGTGTTCTCATT
GTTCAGTTCCCACCTATGAGTGAGAACATGTGGTGTTTGGTTTTCTGTCC
TTGTGATAGTTTGCTCAGAATGATGGTTTCCAGCTTCATCCACGTCCCTA
CAAAGGATATGAACTCATCCTTTTTTATGGCTGCATAGTATTCCATGGTG
TATGTGTGCCACATTTTCTTAATCCAGTCTATCATTGCTGGACATTTGGG
TTGGTTCCAAGTCTTTGCTATTGTGAATAGTGCCACAGTGAACATTCATG
TGCATGTGTCTTTATAGCAGCATGATTTATAATCCTTTGGGTATATACCC
AGTAATGGGATGGCTGGGTCAAATGGTATTTCTAGTTCTAGATCCTTGAG
GAATTGCCACACTGTCTACCACAATGGTTGAATTAGTTTATAGCCCCACC
AACAGTGTAAAAGCATTCCTATTTCTCCACATCCTCTCCAGCACCTGTTG
TTTCGTGACTTTTTAGTGATTGCCATTCTAACTGGCACCACAGTAAATTT
TTATAGATTTTATAAGCAAATTGTATTTACTGTGCAAGAATTGGTTTATT
TTTTAAACCATGTGTTGCAAACATACAATGGTTAATTGTGATATTTGCTC
AGTACAAGATCATCAGATCACTACACAGACTTGAGGTAATTCCACCTAAA

FIG. 4 (20 of 61)

AGCAAAGAGAACTGACCCCACATTAACTGAGAAGTCTTTACTTATTTATT
CCCTATAAACGAGCCAATATGAAGAGAAGGCCTTAATGTGGTTAACTATG
TAATTTTTTTCTGACTTTTTGAAATACTGAGAAGAGCTCATGACTCTCCC
ATCTCCTAATTCTACCTTGGTGGATTTTAGACTGACCACAACTCATGGGT
AAATGAGGGAAGACGAATAAGAAACCTTGCTTTTTTTTCCTCCTTGTTTT
TGGCTGGCTGCAGTGGCTCACACCTGTAATCTCATCACTTTGGGAGGCCA
AGGTGGGAAGATCACTTGAGCTCAGGATTTCAAAACTGGCCTGGGCAACA
TAGTGAGACCCCATCTCTAAAAAAAAAAAAAAAAAAAAAAAAAGGCGACGG
GCGGTGCGTGCCTGTAATCCTACCTACTCAAAAAGCCGAGGTGGAAAGAT
CACTTGAGCATGGGAGGTCAAAGCTGCAGTGAACCTTGATTGCACCACTT
CATTCCAGCCTGGGTGACAAAGCAGGACGCTGCCTCAAAAAAACAAAAAC
AAAACCTTAATTTTTTGGCTATTCTTTTCTGGTAAGAATGGTATAGAGAT
GGGGATGAGGATGGCTATTGTATGAGAGAGCAAACAGGGTCCAAGCAGTG
CTCTGGGCTGTCTAAGGACCAGTAGTCAGCTTAACTTCTCAAATTTCCAG
GGAAGGAGTTCGGAGTGGTAGAATATCCTGGGTATGCCCAAAGCATCACC
TTGCAAATAGCCTGTCATGAATAATTTGTTTCATTTGTTATGACTGGAAA
CTGGCTTTGTGTATGCCAGAGAATGGGGGCAGGAAAGAGAGATTGGTGTC
TTGAGCTCTCTGTGCCTCTGGGGCAGTGATGCTTTTCCTCTCATGTGGAA
GGAGAGCATGACTGAAAAGGTGCACAAATAAGGTGTCTGTGAGAGAAATT
AACCTTCCAGATACAGAGACACAACCTTCCCCAAGAGGTCCTCATTGCTC
TGCCTTTTTTCCTTTTTTTTGCTTGTTCTACCATTAATAACAGAAACTGA
TTATGACCTCAAAAGAGAGGAGAAAGCGACTCTCCCCACCCTAGAGCTAG
TTAACCACCATATCTTCCTAGATCTCAGTTCAAGAGTCACTTCCATCCCC
AATAAAAGCCCTTGAGTGCTGAGCACCTCTCCGTCATAGCATTTGTCCTA
GGGGTTTTTGTACATTTTCTTGTGTGAAACTTGGGTTGACATCTGTATTT
CCGACTAGATTACAGTTTCCTCAAGGGTAGGATGTCTTGCTTGCCATTT
TCAGTTCCAGCATCTAGACAGTACCTCAAGCAAACAAGGCCGAGGGGGGT
GCGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGATGAACATGGTGAAA
CCCCGTCTCTACTAAAAATATAAAAATTAGCCAGGCGTGGTGGCAGGTGC
CTGTAATTCCAGCTACTCAGGAGTCTGAGGTAGGAGAATCGCTTGAACCC
GGGAGGTGGAGGTTGCAGTGACCTGAGATCCACTGCACTCCAGCTTGGGT
GACAGAGCAAGACTTCGTCTCAAAAAAAAAAAAAAAAAAAAGAAAGAGAAA
AGAACATCAAATGAATGAATGAGTGAGATGAATGAGTTAGCAGTGTTGGA
TTTAAGTGTCAGATTCTTCCCAGCTTGACTTTTTTCTTTGGCTTAGTGAT
TTTGAGGTCNCAAGATTTATTTTCCTTTCACAAAGGTGATCACTACCATA
AGATCTTCAGAAAAAGAATGTGGCAAGCCANGTCTCACTAATGCAAATCT
CTATAACAACTGTATCAGTACT
>Contig43
GAGGTGTCATAAATATGGACCGATAGATGAATACAGGTAGGATGGGACAC
AATCTAAGATCCCAGGGGGGGGAGACCACACGCTTGGTTAGGGAGACCCA
AAGTGGACCGTGTGGCCAGAAGAGTCCCGCACTGCACTCTAGTGACAGTG
CAGAAAGTCACTGTGGGAAATCTAGAAGTTTCTACAGGTTGCTATTTCAT
CATAGCACTGTGCAGGCCAACCCTTCCTGCTCCACTGGCTGTTGGGAAAA
GCTTTCTCTTTTCTTCCTAGCCAGGGAGCTCTCAAAGTGTTCCACTCTCT
CACCTCCACCCAGGCGTCCAGGTGTGGAGGACACTTGCCGGCTGCTTGTC
TGCTGACTCATCCCTTGGTTTCACTTGGAAAACCTACCACCAGCTGGCCT
CTTTCCAAGCATCAGCCTCCTCATTTTCTTAATCCCTTAGGTGTGATCTC
ACCTCCACACAGTAGATTGCCTCAAGGCCCAATTCCAATATGAATAAAAA
TGATTATTTTGTCATCTTCCAATCTTCCTTTTAAAATATTATTTTATAAT
TCCCTTTAGGAGGATCACCTAAGTGAAGACTATTTTTACCTAAGAAATGT
TAAAATGTAAAGACATGGTTGTAATCTGGGGATTCCTGTTAAAATGGCTA
GCAGACAGAAGTCAGACGACAGGCTAGAAATGTGTGAAGAGTGGTTGCCT
TTGAAAGGCGGAGTTGGTAATGATTTTCTTCCATTTTTCCATGCTTTCCA
ATTCTCTACAAAGGCCTTAATATTACTTCGATAACCAGGACCTCTGATAA
CCTGCCCCCACCGAGTAAAGACTTAGCTGGGAAAGTCAGCTTCATGTGAG
GTAAAAGGAACCAGGTAATACACAATTCCCACTGCCAACTGTCGGGTGTG
CAGGCCTGAGCTTCCTGCATGTGGGAGGAAAGAGAAAGAAGAGAGAAACT
CCAAGATCCAAGAGATCCAGCAAGAAGGCTGGAGTCTGAGGACGCAGAAA
GCTGAATGGCACAGTTACCACTATTGTGCTGAGGTTCTGTGGCCTCTGGG
TCTCTTGACAACTGGGCAAAGACCCACAGAAAACTATCTCTAGACCCTAC

FIG. 4 (21 of 61)

CTGTGGGAGGGGAAAGT..TTCAGATCA.CTACAGGACAGCCACCTGG..
CTCAAATGGCTTACAGTTCCTTCATCCAGAGGGTCTTCATCTAGTACATA
CCAGGTGCTAAGCCTGGGTGCTGGAGACATGACGGGGAACCCATTTACCA
TGGCTTTGTTACTGTGACATTCACATCTAGGGAAAGCCAGCAAAGGGGAG
GGATCGAGGAGAGCTTGTTAGGCAGAGAAAATACCCAAGGGCAAGGGAGA
AGCCAGCCTGTTCTGAGCACACACAGTGGTTCCATCTAACTGGGCCTCAG
TGCCAGGTTGGACTGGAGATGGGGCTGAGGAGCTGTCACAGAGCATTCTG
GACACAGATGTCACATAGTCCCTTGAGGTTAGGGTCCTTAGGCATGGCAG
CATTGCTTTGAGTTTTTCCTTTTGTAATGTTGCCATTCATGACAATGTGG
AAGATGGGTCCTTGCAGAGAAGGGCAGGGCTGTGAGACCAGTTAGGAGAC
TAAGATGTGAGCCAAGGAAAATGAGGAACACCTGAACACTGGGGCAGGTG
CAGGGCCCAGAGAGAAGCAGATGGCTTCCTGAGGTTTTAAGTAGGTAGAA
TCAAGGCAGCTGGTAAAGATCTTTTATTACATATAAACTGGAATAAGCCA
TCTGCTCCAAGACAAAAGAGTAGGCGGAAAACAATACAAGACAGAAATGG
AATTAGAACAAACCTGGGAGGAATGTGGAATTAGAGTAGAGAGTCCAACA
CTGGCTGCAATCATAAAAATGTAAAACAAACAAAAATTTGCTAGGTGTGC
TTACTTAGAAATAATTAGCTGTCATATTAAGTTCACTTGTGTTATGGCTT
AAATGTGTCCCCCAAAATGTGATGTGTTGGAAACTTGATCCCCAATGCAA
CAGAGTTGAGAGATGGGACCTTTAAAAGGTGATTAGGTCATAAGGGTTCT
GCCCTCATAAATGAATTAATACTGTTATCATGAGAGTAGATTCCTGATAA
AAGGATGATCTCTGCCTCCTCCCCACAGCCCTCTTGTGCATGCTTTCCTG
CCTTTCCACCTTCTGCTATGGGATGACACAGCAAGAAGGCCCTCACCAGA
TGCAGCTCCTTGATCTTGGACTTTCCAGCCTCCAGAACTGTAAGCCAAAC
AAATTTCTGTTTATTATAAATTACCCAGTCTCAGGTATTCTGTTCTAGAA
GCACAAAATGGACTAAGATCATTAGATTATCATTTTTTATCAGACTGTTG
AAGTGAAAAATAAAAATCAAATAAAGAAATTAAGAGAGCTGCATGCAGCA
GCTCATGCCTATAATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATTGC
CTGAGCTCAGGAGTTTCAGACCAGCCTGGGCAACACGGTGAAACCCCTGTT
TCTACTAAAATACAAAAAACTAGGCCGGGCGCGGTGGCTCACGTCTGTAA
TCCCAGCACTTTGGGAGGCCGAGGCGGGTGGATCATGAGGTCAGGAGATC
GAGACCATCCTGGCTAACAAGGTGAAACCCCGTCTCTACTAAAAATACAA
AAAAAATTAGCCGGGCGCGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGG
GAGGCTGAGGCAGGAGAATGGCGTGAAACCCGGGAAGCGGAGCTTGCAGT
GAGCCGAGATTGCGCCACTGCAGTCCGCAGTCCCGCCTGGGCGACAGAGC
GAGACTCCGTCTCAAAAAAAAAAAAAAAAAACTAGCCAGGCATGGTGGTGT
GTGCCTATAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGA
ACCCAGGAGGTGGAGGTTGCAGTGAGCTGAGATCATACCACTGCACTCCA
ATCCAGCCTGGGTGACAAAGCAAGACTACATTTCAAAAAAAAAAAGAAAG
AAAAAGAAAAAAAGAAAAGAAAAAGAAATTAAGAGAAGGGCAGGTATTAA
CCCCAAATATCCCACCATAGGGACACATTAAAGTTTGCTTGGCCACTCCC
CTAGCATAATATATGGAATGTCTTCAAGGACCCTCTGTTGTAAATACAAG
GCCCTGCTGGACTTAATACAACCTGCAGGCTTTGAGATCCCTACTCTGTT
GCCATCTCTCATAGGATTTGCAGACCAAATCCAAATACTTAAAATAGCAA
CACTCACAAACATGCAAATCAGAGCAGAAAAGAAACTTCTAAAAGGCCCT
GAAACTACACTTTATGAGAGAAGACAATAGGGACCTGAGGGTGGTAGAAT
TTTCTCTCTATGCATCTATGTTTCCAGGGCTCACTTTCTCAATAAACTCT
TAAATTGCTTTTAAAGTAAGGGAACAAGCAAACATTACATTTAAGAGAAA
TCAATTTCATAAAGAAGGGGGGATGTCCAGGGTACTTTGCTTCCATGTTT
TGCTTCCATGAATTTGTGTTTAACAGAAGATGCAGAAAAACACACAATTA
TTGCAAAATCAAGGAAATCCACTCTAAACATCCCTTGGTTTCCCAGGCCA
GTGTCACAACTGAAAACACATATTGTGGCTAATTATGTGTCACAAATTAG
AATGACAAGGCAAGAAAAAAAAAAACTCTCTGATTAACTAATAGCAGCCAA
CACAGACAGCCTGTGTAGCTCGACTCTGCTGGTTTATAAAAGGCAGAAGA
AGCAAACGGCTTCTGTGACCGCAACAGGAAGGGCCTCTGCTCTTAATAAA
TAAATAACATTTAAATTATTCTCCCCCATTTGCAAAGCATTTTCCAACTC
ATTATCTCATCTGACCAGGTATTATTGTATCTGACCAAGAACTTGTATAC
NAAATAAAGAATAAAAAATAAATATGGGCCANGCACAGTGGCTCATGCTT
GTAATCCCANCACTTTGGGAGGCCCAGGCGGGTGGATCACTTGAGGTCAG
TAGTTTGAGACCAGTCTGGCCGACATGGCGAAACCCCGTCTCTACTAAAA
ATACAAAAAATTAGCCCGGCATGGTGGCACATGCCTGTAATCCCAACTACT

FIG. 4 (22 of 61)

TGGGAGGCTGAGGCACGAGAATTGCTTGAACTCGAGAGGCGGAGGTTGCA
GTGAGCCGAGACTGCGGCCATTGCCCTCCAGCCTGGGCGATGAGAGCGAA
ACTTCATCGAAAAAACAAAAACAAAACAAAACAAAAAACACCTTAGAAGA
AGCGTTCCTCCTCTTGCTTTCTGAAGACACTCTACGCTGAAACAGTAACT
TTCAATAAACCATCTCTTCTCACCGCACTCTGCGACTTGCCTTGAATTCC
TTTGTGTGCAAGATCCAATAAGCCTCTCTTGCGGTCTGGATGAGAACCCT
TTTTTTGGAATACTCTGACACAACAAATTGCAGAAAGAAAGTCTCACATG
TATAAAATAAGCAAAAAGATTCTCTGGCATCTGAAGAAACAATTTCCTTG
TCAATATTAGTATCACTATAAGTGTAGAACAACCTGTTGTATGATGCTAC
ATAAAGTATATGAATCTGAATACTGTTGGATACAAAGGGAGACTATNNAA
TGTAATACGTCGCCCGAAATGACTACACTGTTGGTGATCTTTCTTTCAAG
AAGCANAATATTGCCTCNAACATCCTGTACATGGTATAAAATTTTA
>Contig44
CCCAGCAAGAACACCAATACAACGGGGGGGGCGTTCTTTGTGAGGGGTGG
GGAGGTCAATTTTTTGGAACCTGCAGCAGGTAACACACAAAACTTCCACA
GCTGCTACCAGCTTTCCAGGAGAGCCTGTGTACCTGGAGAGGAGAAGGCA
AGTGCTTCCGAACTTGACTTGATGTCTTAGATTCTGCAATGCGTAGTCTG
TAGGGACAGGCTGTAGCTTATCCTATAGGCTTGGGCTGGAGTCAGCAAGC
ATCTGGGCTGGCAGAAGATAAAAGATGCAAAGGTGGAGGAAAGCATACGT
GGTCTGGAAGACAGACTTGGTGGGTGGGTGGCTGCTACAACACCCTAGTT
AGAGGTAGAGGGGTAAGTCAGTGTGTCTTCTGCACAGGCCTCTTCCCCAC
CTCATTCTTCATTTCCCATACAGCCTTGCTGAGTTATTCACAAACATCTG
ATTCAACTGGAAGCTGGGTTGAGGATGACCTAAAGGACTAGTGTGATGCC
TGCCCAGGGGTGTGGGCCCATAGTCAGAGTCCAGAGCCTCCTCTCAGCTT
TTAGCACATCTCACCCACATCCTGGGTCCTTAATTAGCAATATGAAAGCA
AGCCAAGTGACAAGATTTTGTCCCTGGGAAGTCCAGAAGCACTCCTTTTC
TCATTTGTATAAGCATAATGATTTGCTTACATAAATAATCATGAAAATTC
AAATCCCTCTCAGAAATCAGGTCATAAAACCATGAAATGCAGCATGTGGG
CAAGAATCACAGGGAAAGGTAGGTCTTGGAAAAGAAAGGATGGCAGGGAG
GAAGAAAGCAGGGTGCCAGGGGCCCTGGGCTGCTGTCCAAGTCAGGTGGC
TCACCGTCTCTGAGAACATTTCACTTTCTGGTAAATGGGGCAGTTGGAGA
TAGAAGGGTTGGGTGAATGCCAAGAGTGAGCACAGCTGAGGTCAGTGCTG
TGCCTGCAGTCCAGGCGGGAGTAGAAATCCTGGGCCCATCTTACCTCCGA
CCTCATTTCCTCCTCTGTAATAATGTGGGGGTGGGGGAAAGTTCTGGTCA
TCAGCCCTAGCATTCCATGGTTCATTTCCTCATCAGTGATGGAAAATCAC
CAAGCAAGAGAACAGGATGGAGAATAACCGGATGGGTGCAATCGGAGGTG
CTATTTCAGGTGAGGTGGCCAGGGAAGGCCCTCTGAAAGGGTGGCTTGAG
CAGGTGGCTGAATGTACAGAAGCTGCCAATCATGAAAGATCTGGGGTACA
GCATGCCAAGCAGAGGAAATGCGAGTGCAAAGGCCCCGAGATTGGATGTG
GGCTTAGCACAAATGTGGCATGGCAAGAAGGCCAGTGTGGCTGAAGCAGC
ATGAACAATGGGTGGAGGGGCTGAGAGGACAGAGGAGCAGGAAAGAGCCA
GGCTTGGGTAGGAGAGGTGTCAACTTGATATATGATGCAAAGCCCTTGGA
GGTTCCCAAACACAAAAGCAATGATCTAATATATGGTTTTAAAAATGCCA
CTCTTGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAG
GCCGAGGCGGGTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAA
CAAGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGCG
GTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT
GGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATTGCGCCACTG
CAGTCCGCAGTCCGGCCTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAA
AAAAAAAAAAAAAAATGCCACTCTTGCTGTGAAAAATTGACCCTGGGGGA
AGGAGGAGTAGAAATGTCAAAAGTGGAAGCAGACCACTCAGGAGGTCAGG
GCAATGGACTGTGCAGGAGAGACTGACATCTTAGACTCGGGCAATAGGAG
AGAAGGTGGTGAGGATTATATTCTGGGCATAAAGGCAACAGAACTAGCTG
ATGGCGTCAACGTAGGAGATGAGGGAAAGAAAGAAATCAAAGGGCATTCA
TAGGTTTGAGGGTTGAGTAACTGGGGATATTTAACAGAAATGGAGAAGTC
TGGGGAAGGGGCAAGTATTGTGGGGGCAGGGGTCAAAAGTTCTGTATTTT
GGCCAAGTTAATTAATATTTGAGATACCTCTTAGGTGTCCAAGTGAAGAT
GTCAAACAGTCAATTGAATACAAAATCTGAATCTTAGCCCAGGATGGTCT
CACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGTGAGAGGATCACTTG
AGGCCAGGAGTTTGTGATCAGCCTGGGCAATAGAGCAAGACCCTGTCTCC

FIG. 4 (23 of 61)

ACACACACACACACACA. AAAAAGTCA..CAGGCATGGTGGCACATGC..
GTAGTCCCAGCTACTCAGGAAGCTGAGGCAGGAGGATCACTTGAGCCCAT
GGTTCAAGGCTGCAGTGAGCTATAATCACATCACTCAATACTACACTCCA
GCCTGGATGACAGAGAGAGACCTCATTTATTAAAATAAAATTTAAAAAAA
TTAATTAAAAATAAATCCAAATCTTTCCTGAGATTCATATTCAGGAGTAA
CTGTCATGTAGAAGGCATATAATGCCATGGGTCACATGATACCATCTAAT
GAATGCCACTGGAAAAGAGAGAATAGCTAAAAACTGAGCACTGGGCACAC
CAGCACAGTGAGGTTGGAAGGAAGAAATGGAGCTAACAAAGGAGACAAAA
GAGGAGTAGCCAGTGAGAAGAGAGAAACATCTGGAGAGAAGAGAGAGCAG
CAAAAGGTGGGTGAAGGAGAATGTGGTCCACCAGGCCCAACAATGCTGAG
CAGTTGAGTAAGTGAGGACCTGGCCACTGAATTTGGCAAGAAAGAGGATG
TCAGCGGCCCTAGAACAAAAGTGAAGAAGAGCTTGAGGACGGAAGCCTGA
CAGGAGTGAACTGAGGAGAGAATGAAAGGTGGAGACATGGAGCCAAGGAG
CACTGAGACTCCCTTGAGTAGTTTTGCTGTAAAATAAAAGTGAGTGCAGA
GACGGGGCAGGGGGACAGAGAAATGCAGGGGTAGCTGGAGGGAGCCACAG
AATCAAAAGAGGGTTTTTGTGTTTAAGATGGTAGTTGTCACATAGCACAT
TAGTAAGTTCATGTGAATCACAACGTAGGTGAGACAGATCACTAATGCAG
GAGTCAAATCCTTGCAGAGCCCCCAGAGGAGGTGATGAAGGGAAGTGATG
GACATCATTCAGATGCAAGTAGGTTAGCAATTCCTGGGGTACAAATAGGA
GGTGACTCCTTTCTGATTGCTCCTGTTTTCTGAATGAGATAGCACATAAA
GTCCACTCAGCCATGTTAGCTGTTGAAGTCCTTGTGGCTGTCATGCCTGT
ACAGACTGGGCTCTCCTCTCCAGCATTTCCTCTCAGACTAAGCTGAGCTG
CACTAGCCGCTGCCACATCCTCTTGGGGCCATCCTCTGCCACACTCCACA
TATTGCTGTGGTTTGCTTGCAACCCCTGGAAGGTCCTACTGGCTGCTCCT
AGAAGAGTCTGGGCGGCATCTCTCCCTTACTCGTTATCACATGGTGCTGT
AAGCAGTGGCCACACACTTTAGCTGGTGGGATGGGCCATCACAGGCAGTA
AATGCGAAAGACTGCTCAGATTTTAAAGCACCCATGAATCAGTAGAATGA
GTTTAGAATTGTAGTCATCAACACACATTAAAAAAAAAAAAAACAGGCAC
TAAAAAAATTAGTTGAGTAGGATAAAGCCATAAAAGATATTAACTACAAC
CCAGATAGGAGGTGCAAAATTGTCCTTACATAAATCAGATGGAAAAAGTT
GAAAGCAGATAAGATAAAATAGGTAAGCATGACATTTAAAAGGTATTCAT
GGGACGTGGTTACAAAACCAACTCACAACTAAAAAGTCTTAGGACCTCTC
GCTGACTTAGGAGCCTGATCCCAACTTTGAGAATGACTCAGTGTGTTACC
CTGTGGCTAGTGTAGACCAATGATCCTGTCTCAGAGTCACTAGCCAACAG
CCCATATCAAGTAATTGAAACTTTGACTCAGAAACCTCAGTGTCAGAACC
TTTGACTTAGGAACCACCTGTAGTGGTTAACTGCAATTTGCACCCCTTAG
TTCAGGGCTTTACAACACCGGGGGCGGGGAGGGGGAAGGCATAGAGCTGA
TGACCTAAAGGAAACCCATTGCAGCAACGCTTTTGTGTTAAGTTTACAAA
TAAGTGTTGTTTTAGAATCCTCCAGGTAATGCCTTTGTTATTTAATGTGT
CTGAGACAATTCTGCACATTAAAGAATATAAAATATTACCTTGTAATTCC
AATTTGAAATGTGTAATTGACATTAGACTTCTATTTTAATTTGAAATGTC
TAAAACAATGTGGTTAAGTTTGTAAAAGGTGTGTGAATTTTGAGTCTGAT
TTACTACATTTTTTTTTAATTTTCTTTTTTTTTGGAGTTTTAGGGATTGC
TTAGATGGCTAGAAAGATCGCTAGGCACATGTCC

>Contig45
GATGTGTGTACGTGTGTGCAAATACCGTGCCTTTTTTGTTTTCTTTTGTT
GAAACAGAGTCTCACTCTGTCGCCCAGGCTAGAATGTAGTGGCGTGATGT
CAGCTCACTGCAACCTCCGCCTCCCAGGTTCCAGTGATTCTCCCGCCTCA
GCCTCCCAAGTAACTGGGATTACAGGCGCCCACCAACACGCCCAGCTAAT
TTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTC
TCTAACTCCTGACCTCGAGATCCACCCACCTCGACCTCCCAAAGTGCTGG
GATTACAGGCATGAGCCACCATGCCTGGCCAATACTGTGCCATTTTATTA
TCAGGGACTTGAGCATCCATGGATTTTGGCATCCATAGGGGTCCTGTAAC
CAATACTGCACAAATACCAAGGGACAACTGTATTCTAAAAAGACCAAAAA
TTAATAAGCAGGACGCTGAAGGTAATTGCCCCAATAAAGTCATGATCCCT
TGCCCAGTGTCTGAACCTCAGCCAGTTTTCATACTCAGGACCTATTGGCT
GCAGAGGTGGTAGGAACCATATGAGAATCCTGCAATATCATGGCAAGTAT
GCACTTTAATGATATCTGCAGTCCTTCCCCAAAAGGACCTTACATTTACC
ATACTGCTATGTCCTGCGTGAGAGGGTAATACTCAGATTTTTTTTTTTTT
TTTTTTTACACAACGTCTTACTGTGTTGCCCACACTGGAGTGCACTGGCT

FIG. 4 (24 of 61)

CGATCTTAGCTCACTGC .CTTCTGTT..TGGGCTCAAGTGATTCTC
GCCTCAGTTTCCTGAGTAGCTGGGATTACAGGCGCCCGCCACCATGCCTG
GCTAATTTTTGTATTTTTAGTAGAGACGGAGTTTTGCCATGTTGGCCAGG
CTGGTCTTGAACTCCTGACCTCATGTGATCCGCTGGCCTCCCAAAGTGCT
GAGATTCCAGCGTGCGCGGCCATACCCGGCCGGGAATTCTTTATATATTC
TGAAAACTAATCCTTTGTGAGACATAAGTGTTGTAAATATTGTATCCCAG
TTTGTGGCATGTATTTTTAATTTTTAATGGTGTCTCTCAATGAAAAAAGC
TTAACACTTAAATGAGGTCAAATTGATCACCTTTTTATTTATGGTTGATT
CCTTTGGTGTCATGTGTAAGGAATGTTGTTCCTTCCTGTCCCAAAGTTGC
AAAGATTTCTTGTGTATTTTGTCCTAAAAGTTTTAAAGTTTTGCTTTTCC
CATCTGTGCACATTTCACATTTGCTACATCTCACTGACTGCTTCCTCTGC
TGCAGAGCAAGCTCCATGAGAGCAGGAGGCATGGGTCCTGCTTCTTGTTG
GTCCCCAGAGCCCTATGTCATGACTAGGACCTGGCAGGGGACTAGTGAGT
AGCTCCTGACTAACTGACTCAATGAATGAATGATTGGATGATTGAACAAA
GTGGTATGGGAGTTCACAGCGAGTAAGAGATGCCTTAGAAGAGATGAAGA
AGGAGATGGTATAGGGTAGTGGTTCTCAATTCTGGGTCCATGGTGGACTC
ACCTGGGGACCCTTAAAATGTACCGTGGAGGATCCCAGCCCAAGAGATTC
TGTATGACTGGTCTAAGATGTGGTCTGGGCACCAGGTGATCCCAGTGTGC
AGCCAGGCCTGAGGCCACTGGATTTGGTGGTAAATGAGGTAACTATCAAG
GGTACAGACGTTGGTTGCCAACAGGCTTGGGCTTGAATTTAAGCTTTGTC
ACTGACTTGCTGTGTCCTCCTGCACTCGTTGAGCCTGTTTTCTCAGCTGA
GAGATGGGTGTGATAACACCTACCTGCTGTAGTTGTTGTGAGAGTTAGAG
GAGATAAGCATGTTCCTGGAATGAAGTGTGTTCTTAATCCATCATAGGTT
TTTTGCTTGTTTGTTTGTTTGTTTGTTTGTTTTTTCCTTTTCAAGAATGA
GGTTGAGCCAGACTTTGACAGCTGGGTGGAAGTGAACATGTGGTGATTG
GGAGAGAAGGGCAGTTTATGTGAAGGGAATGTAATAATTAGAGAGTGGGC
GTGGGAAGACATGCTGGGGAGAGTGAGCAGGCCGGTTAGCCCTGGTAGAG
GGTGCAAGAGAGCAGTGCGGAATCTGCCAGGGAGACAGGTGGGTGACCAG
GGTGCCAAGGGTGTGGCTTTTCCCAGGTTCCCATGGACACAGCCATCCTC
CCAGATGCCCAGCCTAGCTGTGAGTGAGCAAGAGTTCTGGATTGTCTCTC
TCACTCTGTCTTTTTCTCTCATTCCAGAAACAAAGCAGTGACTGGTACTT
AGGAGGAGAATCAGGTCAAGTTGGGAGAAACTTGCTTCTGCTCAGGGGAG
CAGAAGCAAGAATGGAGGCCCCACCCATGCTGGAAGATGATGAGGGTTTT
GGTTCAGGGAGGAGGAATATTGGGGATCTAAAGGGGCCTGGGAGTGGGGC
AGGACCCTGCCTTAGGACAGGTAGAAACATTTTCTATAAAAAATGGGGTG
GAGGTTGATGGTAGGACCAGGCATCTTTAGTTGGCTCCCTGGAGTGTCAA
GCCCTTGAGATGGTCTTTAAAAGCCATGCAGTGGGGTTTGAATCTGGTGT
TCAAGCTCATAGGTTATTAACATAATGACACTTGGAAACTATTTGGGAGA
GCTCAAGTGAGTGGCCTGGAAGTTCTGTGTTGGTGCAGGAGGTGACTTAG
GATGTGCTGCTCCAGACTCATATCTTTGACTGCACACCTGATGCTTCATC
TGGCTATCCTGTAAGCACCTTCAACTTAACATGTCCTACACAGAACTCTT
GATATTCCTGTTCCTCCCCCAGTTCCTCAGTTCTTACCAAATGTTCTTCC
AGTTACCCAATTGCTCAAGTAAAAAATCTAAGTCCTTCTCTTGGATTTCT
GCCTGTTCCCTCAACATCCCACCTATCCATGAGTGTTCTGTGGCCCTGC
CTCTGAAATAAATCCTGCCTTTGTCTCCCAGTTCACTCCAGCCACCCATC
CTGGGGCTGCACCCTCCTCCTTCCAAGCCCTCTCCCTTTCCTTCCTGGTG
CTGCCTGTCATGTCAAGCATATGCATCAGTGCGACCAGGACATTTGAAAT
GCAACCAGTACAATTGGGCGCGGTTATGCCTACCAGTTTTTCTTCCTTAA
ACATTTTATATTTATGTTTGAAAGCATGCCACCTTTCTTCACTTGCCAAC
TTGACAGATTTATTAGTTGACAACATCCGCTGATAGCATCAGTAATAAGT
TAATTGTTTTTGCACATGTAGCTTTAATTATTCTCATTATCATTTATAGG
AGTTATTCTTTGTAAAGGGTAACTGAGTTTTCCAAAACAAACAGAAATTT
GGGGTGGGCCCATGGAGCGTGACTCATGAAATCAGATTCTTAGAAGGACC
TCGGCAAGTCTCTGGGTTGCTGTTAATGAGCCTGGCTGGCTGCCAGGGGT
GTGTCTGCCCTTTATGAGGCCACCACTGTTCAAATGCTTGCCTGCAGCAT
TACTTGCCTAGGTAGTGCTTGTTTCTACTGAACTGTCAGGGATCCAATTC
TTTGTGGTCTAAGTAACAATACTCAGATTCACAAGGAATTGATTAATAAG
CCAGAATGCCAATGTATTACATTTTTGATGAAGACCATATTTACAGTGAT
TGTATCTGCTCAAGCTCAAATTAGGATTAGAGTTCTGACAAATACATATG
TGAGAAGTATGAGGTTAAATACTTGAAATTTGGACTTTTCTAGAAAATCT

FIG. 4 (25 of 61)

GAATGTGATTGCCATTCACATACCTTTCTGGGGATGATGATTCTTGTACT
TTTATTTTAAAAGACATAGAAAACTAACTTAAGAATCAGATTGCTTGGCT
GGGCACAGTGGCTCATGCCTGTAATGCCAGCACTTTGGGAGGCCAAGGTG
AGTGGATTGCTTGAGCTCAGGAGTTTGAGATCAGCCTGGGCAACATGGTG
AAATCCCATCTCTACCAAAAATACAAAAAAAAAAAAAAAAACAACCAAAA
AGAATAAATTAGCTAGGTGTGATGGTGCGTGCTTGTAGTTCCAGCTACTT
GGGAGGATGAGGTGGAAGAATTGCTTGAGCCCAGGAGGTGGAGGTTTCAG
TGAGCTGGGGTTGCAACAGTGTACTCCAGCCTGGGCGATAGAGTGAGACT
CCGTCTCAAAAAAAAAAAAATCAGATTGCTTTATTGCTGGTTTTCTTTCT
AAAACTGAGATTGGGTCCCATCATCCCCTGGCCCCCATTGGTTAATGGTT
CCTCCTTTGTCTATTGAATAAAATACAGATGTCTGCTTTTGGCAACATGG
TTGAATGTAGACACTGCAGGGTCTTCCTGACTCAAAATGATTTAGGCTTA
GATAAAACACATTTGGAAATGCATTTCTGGATTAACACCAAGGAAAGGAG
ATCTCTTTAAATCCCCTTTCTGTTCCCCCCTCCCTACCCCCTCCAATTGG
GCTTAAGTAAGAAGGGTGGTTACCCGCTAGTAAACCCCCTTCGAAGGGGG
TCTTCTCCTCTAAGGAAAACCCTTGTTTTGACATTTGCTTCAATGGGCC
CTTGTATTTTGTTCCTTGCTAAACGGGTGCTAAACCAGGGGCCTCCTCTT

>Contig46
AAGGCTTTTAGAATATTTGCACACTTTAGAAATGGAAATGTTTTTGGGGG
GCGAGTTGTCTTAATATTTCATTTTTCTAGCTTGTGTGACATCCTTTTGA
AAGCAGCAATTCTGGCCTTTGTGAGAGATGGTGAATGCCTGCAGGTGTGT
GGACCAGTGCGTCCCTTCCTTCCTACATGCACGGCCCCCAGCTGGGCCCA
GCAGAGTGCTGTTACAGAATAATTTCCAAGGGCTGTGTCTCTAACCTTTG
GTCTTGTCCCCCATTGCTGTAGATTTGGCCAATTGACTTCATAAGTGCCT
CTTATGAACATAGATGTTGGCAATGGAAGTTGAGGACCAGTCAGTGGTTG
TTTTATTGAACACACAGCGTAAATCCCAACACAATGCTGACCTAAGAGAA
TTCCAGCCACTCTGATTCTCAGTCTCTTTATATCTGAAAGGGTTCTGTTC
CACTTTTTCCCAGATCAAAATGTCCCTGCAGCTACTCAGCAGAGCTGTCG
CAACTTATACGTAGAAGAGGTAACAGTCCACAAACAGAAAGGCACAGGAC
GAGAGTGGTCTGGGTGATGCTTCCTGTGGGGGAAAAGGTGATGAGGGTGC
ATCTGCACACCTATGTTCATAGGTAAGTCTGGGAGGAGGTGACCTCCCCT
TTGGTTGAGGTGCTGAGGCGTCTTGTTAGAATGGCACTATTCCATTTATC
TGATGCAGTCTGTGGGAATTTTGTGGTATGGCCACCACAGGTACCATGCT
GGGAACAATGCCAGATACTGCCTGCTAAGCCACAGCATGAGTCACATGAG
CATTTGTGGGCTTTGGGAACTAAAGTTATTGAACGATAGTTATCTGAAAA
GGAATTTAGGGAAAGGGGACTTTAGTCCAGCGAACAGTTTGCAAACCAGG
GGGAAGGCAGCCTTCAGCGTAAAATGAAGACGTGTGTGCCCCAAATAACA
AAGGGAGAGTTTGTCTTTTAGAGAGTAAATGTCCACGCAAGGTTCCACTT
AGGCAAATGAAAGATGCAAACTTGCTTAGTTCTGATTTGTTTACATTTGC
TGAATTCGGATTGGTCCGTGCAGGCTTTTCTGGGAACTCCAAATACATGT
ATGACCTCTAGTCATACATGGCAAATGGCCGCTTGGCTCTAATTTGAATT
TAGGCCCAGTTAGTCACTCAGGATTAACCTTTTTCAGGGTTCACAGCTCT
GAACAATGGACTTAGACCTGCAGGACATAATCTGTTCCTAACTCTGGGAC
TACCTGTGCCTTTTGACTGTGCCCAGTGAGCAGCTGTGGCTCTGGGCCCA
GACCCACAGGGCGATAAGGCACAGAGGTACGCATGGAGCAGGCTGTCCTT
GCTGAGTGATCATGAAGATACACTTACATAGAGCAGCACTTTTCCTTCCA
GTCTTTGTGATTTAACTCATTAGATCCTTATAACAAGAGTCAGTCCTCTA
TTTAACCCATGAAGCACAGGTGGAGTCCAAGCTTAGTTTGTGAAGGATGA
GCCAAAAGGATTCTTCTCTTGTAGACCTCAAGCTCAGCTCTCTCCATGGG
CCCTGGAGTAGGTGAGAAGGCCTCTGTCTTCCAGAGCCCACTGCCAATCA
TCTACATTTTCTGTTAGCCCAATTCTAGGACATTGCTTTACCAACTGAAG
GGTGAGAACTATCATAAGTTATAAAAATCAATTGAAAAACAAAAAGGTAC
AGAACAGAAAATAAAAGATGAGAATCTATTAAACATAGTGATGTTACTGG
AAAAGGGGGTCTCAAACCAGACCCCAAGAGAGAGTCCTTGGATTTCACAC
AGGAAAGAACTCAAGGTGAGTTGCAGGGTGCGGTGAATTGAGAGAGTTTA
TTGAAAGCTATTCCATTACAAAGTAGAGCATCCTCAGACAGCAAGTGGAG
GAACATGCCATCATTAAATTTTTCTTATATAGGAATCTTGTCTATATAAA
GACTAAACTAAGCTGTGGCTATGTGTGGGTGGGCCGACAGCATGAAAACA
TTTATTCTCCTATTGATTTAAAGAGAACTATCCTTGACATTTTAGTGTGT

TTAAGTACATCAAAGCATAACTATAATTATCTTGAAAGCATATATTTTTA
TAGGGATTGGGACATCTGGCTTTCTGTTGTTGTAGAAGTTTGTCCTTGC
AGGGATTACCAAGCCACTTCCTTAGCTGTAAACATCTTAGGGCCATGGGT
CCTGACTGGCAAGGAATGTGTCTTGCTAGTTTTAAGATGGGCTTGATTTG
AAAATGGTGTCCATCTGGCTCTCCTAGGCTCCTGCTTTCCTAACAGTAAG
GGTAAATGCTATGTTATGAAATGTCATTTCTGCCTTTAGCTTGCAAACTC
TTGATGGTGAAATTCTCCTGTCCGTTTTCAGTGGGGTATTTATTCTGCAT
CCACGTCTTCACAAGGAGCTGAAAACAAATTGGATGGAAGCAACTGGGTT
TTATGGGACACGTTAATGTTTTAATGTCATTTGGTGTGGAATTCAGATGT
CCAAGCAACATTTTACACTACAAATCTGCAACTTTAATAATCACTCAAAG
TACCTGAACCTCAATGCTTTCAGACAGACTTGGTATAAAGCCACCACCTC
TTTCTATTATGGCAGCCCTATCCTGAGGACACAAATTTCTGCAGGGCTTC
TGGCATATCTCTGATTAAACAAATGTCAACAAGGTTAAAACAAATGTCAT
CTCTGATTTGTTTGTTTTAAAGCCTGGATTTACTCATTGAATATTTCACT
CCTACTAGCATGTCTTGTAGTAGTTTTCTTCAGGGACCCTAATTATTGCT
ATTAAAAATATGTGTGCAGCTACATGTTTTTTTTTTATCAATTTGCAATG
AAAACTTTAATTGAATAATCTATTAGTGTTATTATTTGAAAGTGAAATCT
TTTCCTTTTGCTTTCTTGTTCTCACACATAGTGCAGACAGTTTCCACACG
GGCTCATAAAAGGAATGATTCTGCCTTGTGTGAACTTTTTGCCTTTATTG
TTAATTGCACCATTTTGTGACTGGCTTCTTGACCCTGTTGTAACCAAGCT
CATAATGTACATTATTTCTTATTTTGCAGTTGTAGACACTTGAGGAAGTT
CCCATTCTTTGTTTCTTCTTGCTTTTGTTCCCTGTGATAACTTTTTCATG
CAGACATTTTTTTTTTTTTTTTTTTTGAGACCGAGTCTTGCTCTGTCATC
CAGGCTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTCTGCCTCC
CAGGTTCAAGAGATTCTCCTGCTTCAGCCTTTCTAGTAGCTAGGATTGCA
GGCGTGCACTACCACACCCAGCTAAATTTTTCAAATTAGCCACCCCACCT
GGCTAATTTTTGTATTTTTAGTAGAGACAGGGTTTCAACCATGTTGGCCA
GGCTGGTCTCGACCAGGTGATCCACCCGCCTTAGCCTCGCATAGTTGCAG
GTGCTATTCTGAGCTCAGGGCTCTGGCAGCTACAAGCCCAAGATGCGGTC
TCCAACATGTGGCCATTCAATGTCATGGCGCCCTCTACTGGTCCTGGGAA
GCGCAGCTCTGCCAGTAGCTCCAGCAGGGCACAGCTGTTAAGTCGTGATG
TTCTACAGGTGACCAAAGGGCAATCTCTGGACTCCTTAGCCGCTAGGTCC
TCTCTGTAGCAGGACCCAGGAGAAGGCAGGGGCTGAGGATGGCTCTCTTA
GACATTTGTGATGAACCAAACGTGTGCATTCATGAAACTTCTGTGAGCAA
GCAGGTGAGTAGAGTTGGGTTATAAAAAGTCTTAGGGTCTCACTACAGAG
ATGGACTTGCTGTGTAGATGGTGCAGAGCCGCTGAAGAGTTCTACTTGGG
GTAATGGTGTGATTGGGTTTGCGTTTTAGGAAGATTTCTTGGCCAGAATG
AGGCGGGCAACCCAGAGCAGGGAGTGGCCACATGTGGGTGTGCAGTTATG
GGCCACTAATCCAGGTGATAAATGGTGTCTCTGAACTTCAGGTGGGGGTG
CCACATGTCTCCATCTGCTCTGTACCCTTGAGACTGGCCTTATGGGCTGC
CTTAGTGGTCTGTTGTCCTCTATCTCCTGGTTGGGCTCAGGCAATGGGAG
ATCAGAGGGAGGAAAGAGAGCTTGGTTAGAGTGCACCCGCGCCCCTTCAG
GTTGGCAGTGGCCACATTCCCCTATACAGAAGGCCACAGTTTCTGTCAGT
GGCCCTCCCACAGCCCCAGCTTTCTCAGTGGGCCAGCCACCTCCCCATCC
CTTGCTCCTCCTCCTCCAGAGAGGGTTGTGGATTTCCACTGTCAGCAGTG
CCTGGAGCTCCACCATCTCCTGCTGCTTCCTCTGGACCTGCCTGCAGTTT
TATAAATAACCTTTCCTTACATTACCTCTAGCATGCACCTTTTGTGTGTA
TACTCTGCCCCCTGTCAGCACATGACTCATGCCAAAGAGTTTGAATTTTT
TTCTCCAGGCAACGGGAGGTCATTGGAGGATTTTAGACATTGAGAACAGA
TGTGTATTGTGGAAATATCTGTCTGACTGAAGTGACCAGGATGGTCCAAA
AGAGCGAGAATTTGAGGCAAGCAAACCATCAGCAGGCCAGCAGCAGAAAT
CCAGGTCATAAACAGGGAAGCTGAGGCTCACAGGGTTGGATCAGGGAATG
GGAGAGGGAAGCCAAACAATTCCATGAGCATGTCAGTTGCACATATGACT
TGGTAACTATTTTTATTTTTATTTTTATGTTTTGAGACAGAGTCTCGCTC
TGTCACACAGGCCAGAGTGTAGTGGCATGATCACAGCTCTCTGCAACCTC
TGCCTCCTAGGTTCAAACAATTCTCCTGCCTCAACCTTCCAGGTAGCTGG
GACTACAGGTGCGCACCACTACACCCAACTAAGTTGTGTATTTTTAGTAG
AGATGAGCATTCACGCTGTTGCCTTAGACACGG
>Contig47
AATATTGATTATTTGACCAGAAATTCATGCAGCTAACCGTGACCCCTGGC

FIG. 4 (27 of 61)

AAAATAAAATAGTGTAT...GTACGTGCATATACATGCAAAGAAATGAG..
GAAACTAGAAGGATGTCAATCAAATGATAACATGGTCATCTTGGGGTCGG
AGTACATTTGGGGATGAGGGGAGCTGTAAAAGCAGACTTGGACCTTTTCT
TCTACCAGTACCGTGTCATTTGAATTTTGGAAAGAAAAAAAAAAAACTCAG
AAGGAGGAGAAGGAGCAGGAGGAGAAGAAGATGGATCTTAAGTGATTTGC
CCGGGAGCACCTTGAGAAGGTGAGATTCAAGTCTAGGTCTAAGCTTTCTA
ATTCCATGAGTGGGAGTGACCCACGTCCAAGAGGAAGCTCAAAAGGAAGA
TGTTCTCCATCATCTCTTGCTCATCCTAACAGCATGCAAAACCACATCCA
ATGCAGCTCAGAAAACTCCCAAATTGCCAAATTTCATTGGAAACACTTAA
TGCTGTGGTTTCCAATTTCAACTGTAAAGTAGGTATGTATGCCATTGTTA
CCATTAACTTCTCAGAAATGGAGAGAGCTCTCTTTCCGCCTCCTCCCCCT
CTGCTGTGGCTTTGGTGAGACGTGCACTCAGGCTCACCTGTCTCCATGAT
CTCCAGTAAGTACACATGAGCAGAGAGGCCTCAGCTCAGCTCTTCCTGGT
CCCACCAGGGTTGATTCTTTGAGAATTCTAGAATGCCACATCCTAGGCCC
CCCAAAGAAATCCTGCATCTTACCCCCAGAAATATGAATCATAGCAAATT
TCAAATCAACCATCGTTTAATACTCACAGACTGGGCACATCCAAAAACAT
ATTTTCAGTTTTACAACAGTGCCTGGTGCATATCGGCACTATTTGTGGAA
GCAATAAATCGACACGGAGCTGAAACACAAACAAATGCCAAATTGTTTTT
ATAACACCTGATTTTCTTTCTGTTTCTTTATGCAGTTTAGTTTTGTTTTG
CTTAACTCTACCTCAGACCATAGTCTGGTAAACTCACCACCCAGAAGCTC
CCTTGAAATGTGGGTATGCAGCCACTAGGTGGCAGGAGAGAGTTTCCTGC
CTGGAGGGAGGACAGCCACTCTGTCCCCGGGTCAGGCCAGGGCCACCCTG
CTACCTGCAAAATTAGCATGGGGCTTTATGAACCACAGCTTCCTAATAAA
CACAGGATCTGTTTGATAGAGACTCCAAAACACGCCTACCTAGTGATGAA
AGACTCAACTTCAGAAGAAAACCTTCATGGCAAACATCTTCAGAGATGTT
TCCAACTTAAGGTTCTGAACACAGACGCTTCCCCAGAAAGCCATTGTTTC
TCAGCACCTGGGAGCCTTGCTTTGCTTTGCTTACAGACTCGCTGTTCTTA
AATCACTGCCAAGATAACATCTGTCTCTTCTCTTACCCTCTATTTCGATA
TAAGGACTCCTCACTCTTGTTGCTTCCTATTGGCTACCTCTCCACAGGGA
GAAATCGCTGATTTAACAGCAGTCAATATCCCAAATCTGGAACAGGGAAC
AGGGAAGCATTTAAAAATTGGAGAATTTAGGCCGGGCACAGTGGCTCATG
CCTGTAATCTCAGCACTTTGGGAGGTCGACGTGGATGGATCACTTAGGAG
TTCGAGACCAAGCCTGGGCAACATGGCGAAACCTCATCTCTACAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAACCCAAAAATTAGCCGGGCATGGTA
GTGCACACCTGTGAGCCCCAGCTACTCAGGAGGCTGAGGTGGCAAGACTG
CTTGAGCCCTGAGGTCGAGGCTGCAGTGAGCCGAGATCACACCACTGCAC
TTCAGCCTGGGCAACAGAGTGAGACCTTGTCCCAGATAAATAAATTAAAT
TAATTTAATTAGAGGATTTAAGGATTTTCCCTACAGACACCTCCTTATTT
TCTCTGGCCTTTTCTGACTACTCTCCCTAACTCCCTGCTCCTCTGGTCTC
CCAAAACTACTCCAGAAAAAAAAAAAGGGGGGGAGGGACTAAAGGAAAGCC
AGGTGACAGTGCCAGTGTGACAGATGACAAAGCATCTGCCCGAACAAACC
GTAGGTCCCTGAACTTTCTCCAAGACCTGTCTGTGGACTTACCTATGAAA
ACCAGTTTTAGCAAAAACCCTCCTAAGCCAGTTTATCAAGATCCCCTTAT
CCTCAATATCCATCTGATTGGATTCTTCATCCCCCACCATTCCCCAGTGA
TGTCACCAGGCCTTTCTTCAGCAACAGTAGTTAGTGGGTGTAGCCAGGAC
GCCCCCTCACCCCTGATATGCCCTTTTAGTAATTCTTCATCCACAGGTTC
CCACCCTGCTCCTAGGCTATACATTCCCATTTGCCCATGCTGCATTCGGA
ATTGAGCCCAGTTCTATACTGAGGTCTTACTTCACCTCTCGCCATAGTCC
TGAATAAAATTGGTTTTCACATTTAAAAACTGTCCAGCTCTGGTTGTTCC
TTGACACAGGGTAATTTTTATTCCATGTGATAGTTTGCCTTACCTCAGCC
TACACCCCTCAAACCTGCAACTCTATATTCAAGAACCAGACAGCCCTTTC
CAACAGATAGGAAGAGGCTGCCCTGGTGCAAAGGAAGAGGCTCTGGGAGG
AAGGAGAGAACCCGAAGGCTGCCCCCTCCTCTAGACTGAGCTCTGGGATG
GGTGGACGATAAAACCCAGATACGTTTAGACATCTGAGCGTGGAGAGGAC
TTTGCTTTGCTTCCACAGGGACCCCAAGGAAACTGCAAGCCCTCCAGAGA
CTAAAAACAGCAGAACAGCAAGAAATGGCAGCAAAGGTCTGGGCAGAATC
ATCCTATGTGGGCACAGACACAAACAGAGTCCCCTGTGGCCCCAGGAGAG
TTTAAAGAAGATCCAGAGGCTGTCCTATTCCATATCTCAGCAGAGACAGG
CCCGTGAGCCTAAAAGCTGATCATTAGGACAAGAAGGACACGAACTGTCC
TGCAGCGTGAACCGCGTGGAACAAGGCCAATCACCAGACACCAGACCAGC

FIG. 4 (28 of 61)

CAGACACAGCCCCGCAGTTCCCCAAGACCACCACGGACCCATCGCCCCTC
ACCAATAGCTCCAGGCTACATAGACCCCCTCCACTTCATGGATGTCCTCA
GAGCAGAAAGGGGAGGCAGGAGTGGAACCCTGACTTGGTTCAGTTGAAAC
ATAAAATGACTGTACTATTATTGAATTGCTGAAGTTTACGTGAAAGAAAT
GAGATTTAGTTTTTGGCCACAGTGCAAAATAAGAAACGAGGCTTCAACTG
AGATTAAGGTGAGTTATAGGAAAATGTACTCCCTTGAAGGACCTGTGAAG
TGTGTTCGCTATGAGAAAATGACCAGAATCCACGTTCTTAGCTGCGGGAC
TCAGGCTGACTCCTGTTTCTGGAGCTTGCACAAAGGGCAGGGAAATCCCT
GTTTCAGGCACAGTGATTTCAATGTTTAAAAGAAAACAGGTGGCCCCTGG
CAATCATGATAACATGTCATAAGTTTACATCTCTGTGAGGCAGGTAGTGT
AATCCCCATTTTGCAAAGGAGGAAACCGAGGCTGAAAGCAGCTACATGGT
CTCTTCAATGTGGCCCAAATGTTGGAGAACAGAGCTTAACTGAATCAGCA
ATTCTATACTTAGAACTGACTCTCTCTTTATTATATCTCACTACTACCTT
GATATTTGAAATATTCAACTTTTTTCAATCAAAAAATAACAATAATTTAG
GCATAATGACTACTATGTCATTTAATTTCTTGCTGATATTTCAATATCCC
ATGCCAGGAATATTGAAAGCTCAGCTCCTTAAGAGCTGACTATGGCATCA
ACTCCCAACAACCATCCTTCCAGAAATATTTTCCCCTTTCTTTTGTTATA
GAGTGGCACTGCCCTATATGGTGACCACTTGCCACATGTGGCTGTTGAAC
ACTTGAAATTGGCTTGTCAGAATTGCAGTGTAAAGTGTAAAACACATACC
AAATTTCAAAGACATGGCACATAATAAAAAATGTAAAATATCTCATTAAC
AATTTTTATATTGACTGTGTAAGTAACATTTTGAATATATTGGATTAAAT
ACATGGATGATGCCCCAACACCCACAGTCCCTTATCAAGTCTCTACTTCA
CATTTTTGTACTTCTGACTTAGAAATAGCACTGGCGTCTAAGAGCCTATT
AATGTCGTCAATAGGTTCTTGGGAACCACAATTTTAAACAAAATGACATA
TAAGAAAACGAATAACATTGAACAAAATGACATTATTCGAGGACCTGCTG
CATGTTGTTTCACTTAAAGTCAGTGTCCAAGAACCTATCAGTGACATTTA
GTGAGGACTTGCTGTCCTTCCTGTTTACAGGAACCTGGGCAAGTTACTTA
ATTCCTCTAAGCCTGGTTTATATCCCTGCAAAGAGAGAAGGATAATAATC
ACCAGTACTTAGTGATGTCGTAAGGAGAAAATAAAATAATAAATATGAAA
TGGCTGACAGTGTCCTTGTCACACAGAAGATGTGTGATCCACAGTAGCTG
CTATTGTCTGCCTCACTTCACTAGTAATGGTCCAGGGAGGCCTTTAATGT
GCATGGTGCAGTACATTCACATGTTGGACATGGGTGAAGGGAAAGACCAG
GCTCATCTAAACACAATAGGATGCTTGTGGTGTTTTGAGGAGGAATCAAG
GACTAGTTATCCACAGCTGTAACATGCATGGATCAAAAGAGATAAGGCAC
ACAAAAGACTTTGTCAGTAGCAAAGCATTACAAAATGCAGAGACCAGCTG
TGGGTGGTGGTGAGTCAGACCCAGCTTCCCTCTGTGCCTGGCTGAGTGGT
TCTGGGCAAGTCACGCCATCTGTCTTGATGCCCTTCCCCATCTATAGAGA
GGGAGCAACTGAGGCCCCTTCCAATACTGAAGTCCTTTATTTCTGCTACT
TTAGAAATATCCACATTTTTGGTAAATTCAAATGATCCAATGATTCCATT
TCCTAATGTTCAAAACTAGCCCCAGAAACATCTAAATGAATCAAACAAAT
AAAATATTTATTGTGTATGTTTTGATTGCTGAAACTTCTATTTTAGCAAC
ACACACACACACACACAGAACCCATAAGCCTTCATCTTTCCTTGGATAAA
CGAGCCTTCCTGTCTGGCCATTTAAGTCACGATTAAGTAAATGATTTCCA
ACTCGCCTTTTGCAGCAGTTCAGATGGGTCTTTCCTGCGTGGCAGTGGCC
CTCCTGACTTATGATTTCCTGTGTGTCGGCCTGTTACCACTGCAGCTTAA
CTGAGGAAACAAGAACAAAACAGCTTCTGACCCCAAGAGACTGTTGGAGG
CAAAGGCTTCAGTCCCAAGAACCTCACACGTGGGGAGCCCGAGAGCCCAG
CCCTGACCTTTTCTCCAGTAATAACATAAGAAACAACAGGCACTGGCCTT
ATTTTGGATACAAAGAGTGGTGCTTTTCCTTAAATCTTCCTTTAGTCAGG
GCTACCCCTTCATGGACGCCCCAACATCCATGGTTCCTGCTTGAGTCCCT
GCTTCCATATTCCTGCACTTCTCACTTGAAATATCCCTGGAGTACGTTAA
GCAGCCAGGTTTGGAAGTTCTTGCTGTGCAGGCGGGTGTGTGCATGTCCT
CTCTCTCAACAGGACACAAGCTCCCCAAATCAGACGGTATGCCTCCACGC
CCCTTCCCAAGCCTCCCCAGCAGCACCGAGCATGTGAGGGGAGCTGGGGC
CCAGGCCATGATGGGAAGCACTCTCTGCCTAAAGACTAGGGTGATGCGCC
CTCAACTGTGGGAATGAGCCCCAGCTCTGGTGTCTGCCTCGGTTTTTCCT
CCTGGACAATCAACATGAACTCCTCACCCCTCTTATCCACTTTGCATAAA
CTGAAAATAACAAACCCAGGGCTCTTTCTGTCACAGGAAAGGGTTTTTTT
TTATAAAATTAAACAGAGATGATTCAACACACCCAGGATATAACACATGG
GCCATGAATCAAGGGCAGCATTGCTCTGGTCAGCCTGTTGTTTGGGCCCC

FIG. 4 (29 of 61)

CTTGGCAGGGCTCTCCCC GAATCTTCCCCTCTTGACTCCCATCANCACA
GCACTCCANCTTTGTGTTACAGGCGATAAATGGAAAGGGGTAAAT
>Contig48
CATTCTTAATTAGAGAAACGCTCATTAAACTAGACACCCAAATTCTCTGG
GGGGGGATCATTCTTACAAGCATGCCCTTCTCTCTTAAAGAGAGAGCACT
TTTTTCGCAAATAATGCTGCCATGAACATACGGGGTGCATGTATCTTCGT
AATAGAATGATTTCTATTTTGGGGGGTATGTACCCAGCAATAGGATTGCT
GGGTCAAATGGTATTTCTGGTTCTAGATCTTCGAGATCTTCCACACCGTC
TTCCACAATGGTTGAACTAATTCACATTCCTACCAACAGTGTGAAAGCAT
TCCTATTTCTCTGCAACCTCGCCAGCACCTGTTATTTCTTGACTTTTTAA
TAATCGTCATTCTGACTAGCATGAGAGACAGTATCTCGTTGAGGATTTGA
TGTGCATTTTGCTAATGATCAGTGATGTTGAGCTTTTTTTCATATGTTTT
TTGGCTGCAAGAATGTCTTCTTTTGAGAAGTGTCTGTTCATGTCCTTTGC
CCACTTTTTAATGGGGGTTTGTTTTTTCTTGTAAATTTGTTTAAGCTCCT
TATAGACTCACAATAACAAAGACATGGGATCAACCTAAATGTCCATCAAT
GATATAACGGATAAAGAAAATGTGGTACATATATACCATGGAATAGTATG
CAGCCATAAAAAAGAATGGATCATATCCTTTGAAAGGACATGGATGAGC
TGGAAACCATGATCCTCAGCAAACTATGCAAGAACAGAAAACAATTGTTG
CATGCTCTCACTTATAAGTGGGAGCTGAACACTGAGAACACAGGGACACA
GAGAGGGGAACAACACACATTTGGGGCCTGTCAGGGGTGAGGTGGGGGAG
GGAGAGCATTAGGAAAAATAGCTAATGCATGCTGGGCTTAATACCTAGGT
GATGGGTTGACAGGTGCAGCAAATCACTGTGGCACACATTTACCTATGTA
ACAAACCTGCACATCCTGCACACGTACCCCAGGACTTCAAAATAAAGAGA
GACAATACTTCTCCCTTAAGTGTCTACTGTTGCTTTGCAATAAAAAGTTC
CTGCCTTTCACTTCACTCTGACTTGTCCCTGAATTCTTTCTCGTGATGGT
GTCAAGAACGTGGACACTGGCTGGGGCTGGAGACTCACCAGCATCCGGAG
ACCCTCCTGAGCCCTCCAGCAATACAACTTTGACACAAACTATGAAATCA
CAGATCCAAGAAGCTCAAAGAACCCAAGCACAGGAAACATGATGAAACTA
CATGAAGGAACATCAGAATTGAATTGTTCAAAATCAGTGATAAAGAGTAA
ATCTTAAAAGCAACCAGAACAAAATATCCATCATATACGCAGAAATAAAG
ATAAGTATGACAGCAGATTTACAAATAGAAAAAAAAACAAGTGCAGCAAC
AGAAACAAACTATCAATCCATAATTCTATACCTAGTGAAAATTTCTTTCA
AAACAAAGGTGAAATAAAAAAATTATTTTCAGGAATACAAAAGCGAAAAA
ATTAATCACTAGCATTCATCACTGCAAGAAATGTTAAAGGAAGTCCTTTA
GGCAGAAAGAAAATGATACAAGGTGAATATTTGGATCCCTGCAAGGAACT
AAAAAGATCCAGAACTGATAACTTAATGGGTAAACATGTAATTTTCATCA
ACAAGTGAATGAATAAACAAATCATGATATATCCATATGATAGACTACTA
CTTAGAATACAAAAGAAGAACTACTTATGCATGTGATAACATGAATGATA
TTCAAAATTATTATTGAGTGAAAGACACCAGATCAAAACAAAGTACATAC
TGTATGATTCTGTTTATATAAAACTCTATAAATTGCATGCTCTTCTATAG
TGACAGAAAGAAGATCAGTGGCTGCCTGCAGACAGGAAGAGATTACAAAC
GGAAATGAGAATTCCTTAAGAGATGATGGACATGCTCATTACCCATCATA
TGTATACAGCCATAATGGTTTTACAGATACATATATATGTACACGCCAAC
ATAAATATAAGTTATCAAATTACAGTAAGTTCTGACTTAATGTCACTAGG
TTCCTGGAAACTTTGACTTTAAGCAAAATGATGTACAGTGAAACCAATTT
TACCATAGGCTAATTGATATAAAGATGAGTTAGGTTTTTGGTTTTTTTTT
TTTTGACATGAAGTCTCGCTCTATCGCCCAGGCAGGAGAAGAAGAGTTAG
GTTTTACAGCATGTTTCTGGTCACAAGAACATCATCAAACTTGTAAATAA
AGGCACAAAACACTTCTAATATTAAATATCAAAATAAATATGAGTTATAC
AGAATTTAAGAAAGATTAATAAAAACAAGTAAAATCATTATTTATGGGAT
TTTTGGTAATCAGTGAGTTATGTGGTCATAGTGGAAGTGGGTTAAGTCAA
GAAATAAATGTTTGCAAAACAAAAATTTTAAAGATCCTCTCCTACCACCA
CACAAAAAACAAGAAAACACGGTGGGCTCGCTAAGCACTTTTGTACCACT
CGTATCTTATGCGTTTGTATGATTATTGTAAATGCTTTATGATAATTTTT
AGAGACAGGGTCTCACTCTGTGTCTCAGGCTGGAGTGAAGTGGTGCAATC
ATAGCTCACTGCAGTCTCAACCTCCCGGATTCAAGAGATCCTCCCACCTC
AGCCTCCAGTGTAGCTAGGACTACAGTTGTGTGCCACCATGCCCATCTAT
CTTCTTTTTTATTTTTTGTAGAGACAGGGGTTGTGCTTTGTTGCCCAGGC
TAGTCTTCAACTCCTGGGCTCAAGCAATCCTCCTGCCTCAGCCTCCCAAA
ATGCTGGGATTTCGGACATGAGCCAGCAGCACCTTGCCCAGCATTTTATT

TCATAATAATTATAAGTCATTCCTTCATTCATCTTACAACCCACTTGTTC
CAGTTCAGGATCTCGGGTGACCAGAACCTATTAACGTTCACGCACAAGTC
AGAAACCAGCCCTGGACAGGACACCATCCTACCGCAGGGAGAACTTACAC
ACCCACACTCACTCAGACTGGGACCATGCAAAGAACCTAACGTGCACTTT
GGAATGTGTGTTCCATACCCACTAGAACAGCTAAAATTTAAAAGACTGAC
CATACTTGAGTGTTGAACAGGATGTGACACAACTAAATCTTTTAAGCGCT
TCTGCGTAAATGGCACAGCCGCTTTGGAAAACAGTTGGCAGTTTTTCAAG
TTAAATATACCCAAACTCTATGATCCACTTCTCAACAATCAAACAAGAGA
AATAAAAGCAATGTCTACACAAAGATGTATACACAAATGTTCATTGCAGC
CTTAATTATACTAGCCCCAAGTTGAAACAAGCCAAATGTCCATTACCAGA
TGACTGGAACATACAAATTGTGGTATATTGATACAATGAAATACTACTTA
GTAATAAAAAAGAAAGAGCTATTAACATAAGCAACAACATGGATGAATCT
GAAAACAATTATGCTAAGTGAAAACAGCCACACAAAAGTTACATACTGTA
TGATCACATCTACATAAAATTACAGAAAAGGCAAACTAATCTATAGACAG
AAAAGCAGATGAGTGGTTACCTAGGGATGGGGCAGAAGGGACGAAAGGAT
GGATTGCAAAATAGCACAAAAATATTGGAGGGATGACAAATATATTCATT
ATCTTGATTGTGGGGATAGTTTAATGGGTATATATAGAGATCAAAGCTCA
TCTAATTATACACTTTAAATATATGTATTTCATTGTGCATCAGTTATTCA
TCAACAAGACTATAAAATAATATATGCCTACATACATTTTTAAATATTCA
AAATCTCACAGTTATATACATAAATGCAACTGAATATGTATTCAGATGTT
TTAACAAGCAGAAAGGACTGATTAAACTCATGACAGCGGCTGTTTCTGGG
AAGGGTGTAGGAGACAAGAGATGGAAAAGAGGATGAGAGCCAGAAGAGAC
CCTTGTAATGTTTCCTTTCTTTTAGTAAAAATATATTGACAGTTAAAGCT
GAGAGGTGAGAATAATAGTCTCATGGCTTTTGTGTCCTTAAAATTTCACA
AACTAAGTGAAATGGGAGAAAGCAAAAAAATAAACTTAAATAAATGTTAT
ATTGCCCAAAAAGAGATTTAAAATGGAGGTTAGACACATGAGACTTACGT
TCTCAAAAAAGTAGAATCTGCAGGGAAGTTTAACAACTATAAAGAATTAA
AATCTAGCTTCTACCAGCCCAAAGCCTAAAATGTTCTGCTTTATTCTTCC
TTATTATAATTCATAGGTAATATATTTTATGTTTGCAAATGAATGCAGTG
ATATTAGATCTCTAAGAGGTGCTAAAAATGAAAAGTACATATTCCAATTT
TTCCCAATTTTCCTTCTCTTTCCATGAATGAAAAATATACATATTTGATG
ATTTCCAAGTTTATACAACCGATCTTTCTCTTAGTTTTCTCTTACCAAAT
TCCCTCCCTCACTCAGCCACCAGCCAGTCCAACTGTGCTACCTGCACAGC
AGCCCTCATACCATCCACACTCTCATCAGGATCCTGCCTGACCTGCGAGG
AGCAGCAGCAAGAAGGAGACAGAACCTCCACGCTGAGCATCTCAGGGCTT
TCTCAGAGACTCCAGAGGACCCTGATAGGGACAGAGCCTGGCCAGCAATC
CATGCTGCCAGCTGTATGATTGTGGGCATGTAAATTCTCAACTGAAAATG
GGTGTAATAATAACATGTTCTTCCCAGAATGAGCTTTATGAAGATCATAT
AGCTGTTTGGAACTCAGACAAGCACTGGTAGGAATACAAACAGGGGAGCC
AACAGCCTATAAATAATACTTTAAGAAAGGGCATGAATGTAATTACTTAG
GAACAAAAGGCAAAGTGGAGAGATGCCTAGGACTGAGCTGGACAAGCTGC
ACCCTTTAGTGGCTCAGCCCATGGGCTGACAAGGAAAATGGAGGAGCTAC
CAAAGAAGGTGGAAGGATTCTGGGAGAGTGGCCCTCACCCTGCCCAGGGC
AGGGCTCAGTGGGAGAGAGGGAGATCTGTTATAAATGCTGCCAGGAGGTC
GAGTCATGTGAGAATGTCCATGTGAAAACATCCACTGTGTGTATCTAAAG
AGAGTGGCTGTAAAACAGGTCAGGGTCAAAGGTCTTATTGTCTCAGATGT
TATCTGCATGCATTGTCTCACGACCAAGAAAACTAAGGAGCATGGACACA
AAGGGTTAGGTTGAAGCAAAAATTTAATAAGTGAAAGAAGAAGGCTCTCT
GCAGTGGAGAGGGGAGTCTGAGTGGGTTGCCACTTTGACAGCTGAATCCA
AAAGCTTTTATAAGAAACTCTTCTCATATCTGCAGCTGTTTGAGTAACTT
CTCTTACCTATAAAACTGTCTGTATAACTCTCCCTTATCTATGCAGCTGT
GGGATGTCTCCAGGTAAGCATAAAGTGTAGCTTCTCTTGTTTGTATAACT
GTGGGTTTGTTTTAGGCAAGCCCCCATCCCCTCCCTGTGTAAGCTCCCAT
GGAGCCCACCATGTGCATATCTGAGAAGTGGAGGAAGCTTTCTCTGGGAG
CTCACTGATCGTACAAAGAACAAGAGGCTTCTGTGCCGCTTATCTATTCA
GGTGCAGCCTGAGTTTTCCCCAGGCTGCTCTATTTTTGCCTGTAGCTATG
ATTTTCAGGCAGGCTGCTTCTCTGAAGACTAGCCTTAACTGTCTACCTA
TCAGATTTTTCCTTTTCTTCTCCCTCAGCTGGTTCCCCTCACCAAGGCTG
AGCAAGTGAAAAGGAGGGCACAGGGCAGGCCAGTAGTGAGCAGCAACAAG
GAACTAAGACAGCAGAAACCACTCTTCACACCTGGGTTGAAAGGGGTGGG

FIG. 4 (31 of 61)

GAGCCAGGACTACAGCTCAGGTAAGAACATAGGTAAAGAGATACTGTTGT
TGTGTTGTTTTTAACTATGAGAAGCATTGAGCTTTAAATTTCTACAGGAA
GGATCCAGTTCAGACAGGAGCACCCAATATTCAGAAGAGAAGAACATGGT
GTAAAGGTCCTGGGAAGGCTGAGAGGATTGGACTCAGAATCCAGAGCAG
AAGCCGTCTGTGAACAGAAGAAGGACCTCCCCCAGTGTAGCAAGAGGGAG
GGAGGAGGGACAGATGCCAAGATGGTTCAGGAAGAAGGTTTGGTGGTAAA
TGTGAGGCTGTGCTCACCTGCTGGCTTCAATTTTCTCTTTAAAATGTCAG
ATGGAATCATTTGATGAAGGCCATGCCATGCAATGAAATGGCAGTCTGAG
GCATGGAGCAGCTCCAGCTTAGCCCGTGTTTAGGGTAATTATGGCTCCAA
CCCAGGAGATGAATATGACTAGGGAAAGTGAAGTCCAAAAACAAATGGTC
TCAAGTTGACTGTGAGTCTTCTGGGAGGCTGAGACGACAGGTGGGGTTGA
CAAGGGAAGGGGAACCCACCTGCTGAAAAACATCAGGCTGTTGGCTGGGG
GAGGGGTGAGGCCTGTGTTGTAGAGATGGATGGATGCCTAAAGTTGGGTA
AAGGTTTCAACTCTACCCTCTGCTGGGTGTGGAAATAAACAAAGACCACC
CAAATGAGAACAAACAAAGACTATTTATCCAGAGCTTGCTCTGACAAGGG
AGTCGGCAACCATCACTTGCTTGGCAGAGACTCAGAAGTAAGCAGGGGAG
AAAGCCTCATAGCAGAAAGAAGGGAAGTCTTCATGTATGCCCTGAGTGGC
AGCTGTAGATGTGGGTGAGTTGCAGGTGGCTAACTAGAAATGGGGGACTC
CTGTGTGATTGATTAGGAGCATGTTTGGCTTTCTCTGGTTGGTCCTACAT
TGGAAGAGGGAACAAAAAATTTAGGGCAGTTGTCAGTTATTAATCAAGTG
TTGGCCATTTTTGACTGACTGTTACAGGAGTGACTGGCTCCCTGGATTGT
TTGCTAGAAATAGTGGTCTTCACTTCCTGCAAGTCTGACTTTCTGGTAAT
AGGCTTCCTGGGTTGGCTATTGTGGATAATAAGTGGGTTTCCTGAGCTGA
TTTCTGCAGATTGTGGATCAGAGTTATTTTATATAAACAGTCTGACCATT
TTCCACTGGCATATTCCATCTTCCAAGAGCTGGCCAAGCTGCTGTCTTAT
CTGTCTCCCCCAGCCCCTCCACTCTGGCTGTGAAAATACAAGCCACTAGG
TGAGGAATGGGGACAATTGAAGACTGAAAGCTTTTCTTTGCTGGGTTCGC
AGAGCTGAGGAAAGAAATGACAACATCCAAGTGTCTGCCCTGGGCCAGTT
TTAGGACTGTAGTGGTAATGCAAGGACTGTGTGAGTTTATATTTTCATTT
GTCTCTCTAACTAAGGTGGAAAAAAAAAAACAGAAAATTGTCTGTCTGCA
GTCTCTGCAAAAGTCTAACACTGTGCTTCCCAACATTGCAGCCATTAGCC
ACAGGTGAGTATCAAGCACTTTAAATGAGACTGGTCCAAACTGAGATGTG
CTCTGAGAATAAAACACACAGCAGATTTCAAAGACCTAGTACATGCCCTG
ATTTCAAGCTATATTACAAAGCTGTGGTAATCAAAACAGTATGGCATTGG
GAAAAAAATAGACACATTGGTCAATGTGACAGAATAGAGAGCCCAGAAAT
AAACCCGTGCATGTATAGTCAACTAATCTTTGACAAGAGTACCAAGAATA
CACAATGGGGAAAGTCTCTTCAATAAGTGGTGTTGGGAAAACTAGATATC
CACATGCAAAAGAAAGAAATTAGACCCTTGTATTACACAAAATCTAAAAT
TAATTCAAAATAGAAAAAGACTTACATGTAAGATCTAAAACCATAAAACT
CCTAGAAGAAAACATAGGGAAAGAGCTCCTTGACACTGGCATTAGCAGTA
ATTTTTCAGATATAACATCAAAAGTACAGGCAATGAAAGCAAAAACAAGT
GAGAGTATATCAAACTAAAAAGTTTCTGCACAGCATAAACAATCAACAGA
GTAAAGACATGACGTATGGAATGAGAGAAAATATTGACATCTGACAAAGG
GTTAATATCCAAAATATATAAGTAATTCACACAACTCAGTAACAAAAGCC
AAATAACCTGACTTTTTTTAAAATGGGCAAAGTACCTGAATAGGTATTC
CTCAAAAGAAGACATACAAATGGCCAAGAGATGTATGAAAAGCTGCTTAA
CATAACTAATCATCAGAGAAATACACAAATCAAAACAAGATATCATCTCA
CACCTGTTAGAATGGCTATTATTAAAAAATGAGATAAGTGTTGGCCAGGT
GTGGAGGAAAGGAAACCCTTGTACATTATTCATAGGAATGTAAATTAGTA
CAGCCATTATGGAGAACAGTATGGAGATTCCCTAACAAAATTAAAAATAG
AATTACCATATGACCCAGCAATTCCACTTCAAGGAATACATTCAAATACT
ATCAGTATCTCAATAAGATACTTGCACTCCTATGTTCGTTGCAGCGTTAT
TCACCATAGCCAAGATACAGAAACAAGTTAAATGTCCATCAACAGATAAA
TGGATAAAGAAAATCAGGTACATATATATATACAATGGAATATTATTCAG
CAAAATCCTGACATCTGAGATAACCTGGATAAACCTGGAGGACATTATGC
TAAGTAAAATCAAAGCCTGACACAGAAAGACAAATACCACATAATCTCAC
TTACATATGAAATATGAAAATGTTAATTTTATGGAAACAGAGTAGAATGG
TAGTTGCCAGAGCCTGAGAGTAGAGAAAATGAGATGCTTGTCAAATCAAA
TCATCACATTGAATATATATAATCTATTTGTCAATTAAATATTTTAAGAA
TAAAAAATACCTGGCACCAAAAAAAGAATGCAAAATGTCTCAACAATGTT

FIG. 4 (32 of 61)

ATATGTATTGCATTTTGAAGTGATAATAATTTGAATATTAGGTTAAATAA
AATATATTTGAAAAATTAACTTCACCTATTTCTTTCCATTTTTGTTAACA
TAGGTACAAAAAAAAATTAAAATTACCTATGTGGCTCATGTAGGTGGCTC
ACATTATACTTTGATGACACTATACAGGCTGGTGACCATATATCTCTTAG
ACTAGTCTAAGTGATTTAACAGTGGTTCCAGAAAGATCCAGGTTTAACAC
CAATGAAAGGGCCAGCTGGCTTAGCCCAGCTTGTGTGGGAAATGTTGGGG
AGTGGTTTAAGACAGGGAAAAGCAAAACTTTTGATGCTATTGACTTTTTG
AAAAATCTTTTGTGGCTGAAAAAACCAAAACATTATT
>Contig49
GCTCGAGTGTGTCTCTAAAGCCTTTCCCCCATTGGCTCCACTATACGCAC
TCTCCTGGTTTCCTCCCCTCTAGCCGCTGTCTTTGGTCTCCTTTCTGATT
TTGCTGCGTCCTCTGTCCCCTGAATGATTGCTTCTCCACTACGGGGTGAT
TTTGCTCCCCAGGGGACATTTGGCAATATCTGGAGAGGTCTATGGTTGTG
TTTGAGGGTGTTGCTACTGCCATCTAGTGGGGAGAGGCTAAAGATGCTGT
TAATGCCCAGGACAGTCCCCATAACACAGAATTATTCAGCTCAAAATATC
CATGGTGCCAAGATCAAGAAACCCTGCTCAAATATTAGCATGTGCTGAAG
GCCCTTCTCTTTCCTTTAGCAATATCTGCCTCCTTAGGGATCTTTTCTAG
TCTCAGTGGTTTAACATTTAAAATCCCAAATTAGGCAATAAATTGGGCCC
CAAACTTCGTTAGTATAAAATGTAGAACTGTGTTATTAGAAGGCTAATAA
AATGACCTGGTGAGCATCTGCAGCTAGCCTCTGAGCAATTCTGGGGACCA
CGTGCAAGATAAATCCATCTGTTCCCTCTCTGTAATGTGGCGCTACCTTG
TGGCCGATTTTTCCTCGGGTTAAATATCTCTGGGGATGCAACTTGTCGTG
GTTAATGGCTGTGTGAGGCCAGCGCGTGGTGATAAAGGAATCAATCAAGA
CAATATTGAATTTAGAAAGGCAGATTTATTTAGAGAAAAGGAGAGATACG
TTGCAAGGGAGCAATGGGCAATACAGCAGAGGGAAGGCTGTCTGCAAAGA
GGCAAGGGCTACGTATGACGTAGGGCTGCTTAGGCTGAATGCTTGCAGAC
AAGATGCTTGCGTGCAGGTGGGCTGTGAGCTGAGTGCTTGGGTGCTAGTG
AGCCATTGGCAGCTGACCCTATTTCTTGGAACATTCGCTCCCTGCAAGCA
TTTTAATGTTAAACCGCCAGGTCAGTTTGAATTTTCTTTTTTCTTTTTTT
TTTTTTTTTTTTGCCTTTAGTAGGACCTGCCGTTGTGAGACTATCTGAGG
TAAATTAGACACCCTCCTGGTTTAAGTCACCGCTCCAGTGACTAGGCAGG
GAGCTCTTCCTTGAAGAGGGTGTGGGCAGTGGGTACTTTGCATGTTGTCC
ACACCAGGCGAGCTGCTGCTTCAGGGCCTTTGCATTTGCTCTTTTCTTTG
CCCAAAATGCACTTCTCTCACTGTTCACATGATTTTTCTCCCTCTTTTCC
TTTTAGTCTTTGCTTAAATATCACCTTCTAGGGAGGCCTTCCCACACCAC
CTCTTCAAGATTTGAGGGTATGCACCCCCCACCCCTAGCCTTCTTATCCCT
CTCCACTGCTTTCTTCTCAAAGCACTTGTTACGTTCAAATAAAATAGATT
AGTTACTTTATAGTTCTAATTTTACTATTTTTTGTTTACTTCATCAATAC
CCATGTAATCTCTGGAAGGAACGTTTCTTTTTGTAGTGTATTTCTAGCAC
CTAGAACAGTACTTGGCACATGGCAGGTGTTCAAAAGTATTTGTTGATTA
TTTTCTCAAAGGGCATGGAGTCTTAGAAGTTTGAGAACACAGTTCTAAGC
ACAGCTGTTTAGAGACTATGGATGATGCTAATGGCTGTATTCCCAGTAGG
TGGGGCAATTCTCAAATTGACCTGGAATCCTTGAGATCTGGGGACAGTCA
CCAAGCACTGGGCTCTGTGGGGAGAGATGTGCTGGTTTTTAGAGAGGAGA
ATAGCATCCTGGGGGACTTGGCCCCAGGGCTTTCCTGTCCCAATCTCTTC
CCAACTGAGTCCCAGAGGCAGGAGGCCTTGTCTGTAGCTGGTCAGTCCTG
TAACTGTTTCCCTCCCATCTACACAGATGCAAAGAAGGCTGAGAAAAGCA
AGCTGTCAGGTGAGCAGGGGCCCTGACTCCTCCCCAGAAGGCACTCAGAA
CTTCCATAGGGCAACTGGAAAGAAGGTTCTACTTCCTCACCGGCAGCTGT
TGCTGGGGAAAAAACCAGCCTCAGGCCCTACCCTGTGCTGAGAACCTGAA
TCCAGTATCAGGTTCTCCAACAAACTTGGATCCAGCTGACCCTCACAAGG
GGTCAGATGCAACCTTGTAGCATATGGAAAATGGCAGCAAGGTCCTTGTG
TGGACTATGCCTAGAATCTAAATTAAGACAAGGCCTCAGAGGGGCTAAGT
GACATCTGTCTCCAAAGTTTCACAGCTAGTGTGTGACTAAATCTTGATTC
CACCCTCTCAGGTTTTACCATAATCCCAAAAAAGGTTGAAACAAGAAAAG
TTATCTTTGGGCAATTACCTCTTTCTGTTCCTTGCTTTACCTACTAATGT
TCTAGGCTCACCCTCTGGTCTGCAATCTCACTGAACTGACAGATCCCTCA
TGGCCTAAAGGGTTTTCACACTGGGTTGACTAGGCTCTCCCATTGCCTGT
CCTACTGTCTAAGGCACCTCCTGGGTAGGGTGCCCAGCGTCATTCTGATG
CTGCCTGACTTTCCTTCCAGCTACTTTGAAACTTGGTATCCATGGCAGA

FIG. 4 (33 of 61)

GGCTTAAAGGGCATGTTC-AGGTACTTTTATTTCCAAATTCCCCAGTGGC
ATCAAGGAAATCAGCATCTCTGGATAGCTCTACTAAGGCTTAGTTCTCAT
TGTCCAATCTAGCTCCTGGGTCATGGGAGGCATTCAGGAAATATTTGAGT
GTAAGAGTGAGTTGCTTTACCTCCAGAATATCCTTCCAATGGCTCTGAAG
CAGGCTGTGGAGTCCTGCTGGCTGATCACAGTTCACAGGTGGCTCCCAAA
CCTGTGGTCTACATCCATCCTTTGTCAGTGTCACTGCCATTGTCCCACAA
ATGTCATTTGGGCCTAGCCCCTGGGATAGTAATCAGTCTTTACATAGATA
TACATTGTGCTTTACATCCACAGTAATTCTGAGTGGACCTTAAAATAAAT
TCCATGTCAGGTCTCACCAGCCCATGGGTTACAGATGGGGTTACCTTTCA
GCCTTGTAAGGTGCCCCGTCTTTGAGTGTAGACATGGACTCACAACGAGT
CCACTCCTGCTGTTCCTCTGCTCTTGCTGAGGCTTCTGCTGCTGCTGCTG
CTGCTTTGCAGAGGCTGGCCAGCTGTGGTGCCTGAGGCACCTGTGTCTTC
ACAGCACCAACTTGCATGGTGGCCACGGTGTAGTTGGAAAGGGATGCTTA
GATGGGAGGCCAATGGGAGCTGCTTCAGGAGGCAAATCCAAGTCACAGAG
ATCGAGTCACCGAGAGCATAGTAAACTCAAAATCCCTTCTTCTGCTTAAT
AACTGAGATGCTGTCACTGGGTTAACCTCACCAAGCCTTGTTTTGTCTTC
ACTTAGAGTGATTTCTGTCTTAGAAGGCTCCTCATATCCTTCTGGGGAAG
GCTTCTAGTGAGTCCACAGATAGCTGGACCAGGCATGTCCAGAAATAATC
TGATTCTCACATTTGAGTTAGCCAGCGTTCCCAGCTATATCCCCATTTTG
TGTCTATATAAGTTACCAAAGCCCACAAGGATATTAGGTGGCTCCTTAGT
TTGCTTTATGATTATGCCTTGTGTGTGTGTGTGTGTGAGTGTGTACGCCT
ATGAGGATTCCTTCTCTCCCGTTCTTGCTATGGCTTCTCTTCCCCACTGA
TGGGCTGTAGTTCCCTGTCCTTTTGACTTTGGGCTTAGTCATGTGACTTT
TTTGCCAAGGGAATGTGGGCAGAAGTAACTGGGAGCCAGTCCCAAGCTAA
GGCCTTGGGAAGCATGGTGAGCCTATGCCAGCTCCCTCAGAACTCCTTCC
CTTGGCCATGAAGAGAGAATAACCTGGATTGTACCTTCAGCCCATGTCCT
AGAATACAAACATGGAGAATAATGAACTTGACTCAAAGGCTGAAGGGCAG
CTGAGCCCACATGAGGTCAATTGAACTGCAGCTACCTACAGACCTGAAAG
TGAAATAAACATGTATAAGTCTCTGACGTTTGGGGTTTGTTTACATAGCA
TTATTGTAGCAGAAACTTAAATAATACTGGGGGCTAAATATAGTGGACCA
GTGACAGCACAGAATGGTAAAATGGAGTGATTGTTACTTACATCACAACC
CTTCATCTCTGTTGATGGACACTAAAATCAAAGTGGCAATTACTCAGAGT
TGGGAGTCATTGAGTTGCATCATTGTTGTTTAGAATCATTGACAGTTTGA
GCTCTAAGTGATTACAGAGATGGTTTCCTCAGCTACAGGTAAATAAACAA
AGGCACAGAGAAGTAAAGTGACTTCTAGAGGGCTTCATTGATATTTAGCA
GCAGAATCAGAGCTAAACAATGAGTCTCTCATCTCCAGCCTTTCTATTCT
TGTTTCCTAGGTTGGGATTTTGGGAAATAGTGCAGAGAGATTAGCAGTAG
TGACATGGAACAATGTGAGCCTCAGCTTCCATCCCTGAGGCTGCCTTCAT
CTGCCAGGGAAATGTCTCTGTGTGCAGCCTTGCCCTCTGCACACAGTGTG
TATGGCCACCTGAATAAGTGTCCTTTCATAGCGACTAATGGATTGAAATG
GGTGCTAGAGCAGTGCTTCTAAAAACTCCATGTATTAATCATCTAGGGGT
CTTACCAAAAACGCATGCAGATTCTGATTCAGTAGGTCTGGAGTGGGGCT
TGACATTCTGCACTTGTAACACATGGACCACACTTTGAGTAGCAATGTAT
TAGATCATTCCAGTGGAAACATGTATGAGTGATGGAATGAACAGATATAA
TTAATCCAGGTCTGGTAAGTGAGGTACTGATACATATTAAGTTGAAGTGA
ATTTCACATCAAAAATAATGGTTACACAGTGACTTTTACTGCCCCCAAAT
TCTTTCCTTTTGAGTGGTTTCAAAGTGAACTGAGCCAGCCAGGTTAAGTC
CCTGGTTTAGTGTGTGATTAGAAGATTTGATCCAGCTTTCTCCTCCTTCT
AATTCTTTAAATATGCAATGGCCTTCTAGAAACTTGTCTCTCAGGCTCCC
CATGAGCCACCTGTCTTAATATCTTCCCCCCCAGGACATTTCCTGGGTCA
AGGAAGGAATCAGGGACTAGGAAAAGTAGAAAGGTTGCCTGACAGTGAGA
AACTTTTTGCACTCCTATTTGTTCAATTCTAAAATGTGGGTATTGTTGGG
GCTTCTAATTGGAATCTAACCTGAAATTCAGGCATGTCTAGCTATATATG
ACCAAGAATTAGGATGAGTTCACTAGAAGCCTATTTTCAGGAGAGCGGTC
AGTTAAATTGAAGTTTATGGGTTTATGGTAATGGGTTGGGGAGTTTACTT
CATTAGCAATAGCAACGTTTTTGAATCAGAGAAGTGATTTTGAACACACT
GTACATAGTTTTCTCACTTAGATTTATCTCTGGGTCAACCCTTGTTGGAC
CTATATTAGAATCATTTAGTGAAGAAAAGGTGGGTGTCATTAGGAAAAGA
GCCATTTATTCAAATGTTCTGTTTGACATTAGGGCACTGGCAAGACTACA
GAATCAATAGATATTTAAAAACAGCCAGGTGCGGTGGCTCACGCCTGTAA

FIG. 4 (34 of 61

TCCCAGCGTGATTTGGC .TTACTTTGGGAGGCTGAAGCGGGTGGATTC
TGAGCTCAGGAATTCAAGACCAGCCTGGTCAACACGGTGAAACCCTATCT
CTACTAAAATACAAAAAATTAGCCGGGCATGGTGGCAGGCGCCTATAATC
CCAGCTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCG
GATGTTGTCATGAGCTGAGATCGCGCCATTGCACTCAAGCCAGGGCAAGA
ATAACAAGACTCTGTCTCACAACAAACAAGCGAACATACGAAACAAACGT
AACATCCAAACTAGCAGGTACATGCCGTGCCAGTCATGACCCATGGTCAT
AAGATGTCTACAGCTCAGGAAGCAGCTGCACAATGCCTGCATAGACAAAC
TCTTATGAAAGCAGAATGTCCTGATGTCTCCATAACACATAACAGTGTAT
GCTTTTATTATGGTCATACTCTAGCTGTGATGTACCTACGCTCTAATATG
CCAACGATAGTTTTCTTTAAATCATCAACATAATAAATGTCATGCTGTCA
GTCCCCCACATGTAGACATAACTTAGCTGGTACATGGATAAGAAACCTAT
ATTAGATAACCTTAGGCCAGGTGTGGTGGCTCATGCCTGTAATCCCAGCA
CTTTGGGGAGGCCGAAGCGGGTGGATCACGAGGTCAGGAGATCGAGACCA
CCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAAAAAA
TTAACCGGGCATGGTGGCAGGCACCTGTGGTCCCAGCTACTCAGGAAGCT
GAGGCGGGAGAATGGCGTGAACCCAGGAGGCGGAGGTTGCAGTAAGCCGA
GATCACACCACTGCACTCCAGCCTGGGGGACAGAGCGCAAGATTTCGTCT
CCCAACCCAAAAANCNANNNNAAATTTGCACCCAAATCTGACTAATTCCA
GAGCCAATTCCAATTTAGAATCGTTATATCTCCCTGGTGAACTGAAGCTT
TTATCTTTAAGGAGACACACTCTTTATGTCTACCAATGCTTATTGCCTTA
AAGTCCACTTTGTCAGATACAGCTGCTTTCTTTTAATTAGTTTTTGTGTG
GTATATCTCTTTCCATCCTTTTTCTTTCAGCCTTCTCCATTCTTACATTT
TAGATATATTTCTTTTTTCTTTTTTTTTTGAGAGAGAGTCTCACTCTCTC
GCCCAGGCTGGAGTAGTGCAATGGCGCGATCTTAGCTCACTGCAACCTCC
ACCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGG
ATTACAGGAGCCCACCACCAAGCCCAGCTAATTTGTTGTATTTTTAGAAG
AGATGAGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCA
GGTCATCCACCCACCTCGGCTTTCCCAAAGTGTTGGTATTACAGGCGCGA
GCCACCATGCCCAGCTGATTTTAGCTGTATCTCAAAAACAGCATGGGTTC
TGTTTGCTTTCCTTATTCAGCTTTATAATGTAAATCATTTACATCAAACA
TCTAATACACCATGGACTGTAAAACACAGCCATATTTTATGTATGAATTA
AAAAAAAAAACACCACCAATTAGTTCCTGAGACACACACCTTAACAATAT
CTCTGTGATGTGCATAAATCAATCACATCAGTTTCTCTGCACCTCAAAAT
TTCTTTCCTCAATTCTCAGAGATATGGCAATTTCTCTGGTTTTACATTCC
CAGAAGCAAAGAAAAAGTACACAGCTTCTTCAAGTCATGAGTAGCTTCTT
TTTTATAGCTCTTGGTGTTTGCAAAAAAGATTGGAATTGCTTCACTAATA
CTAAATTTTCATTCTGCTGCTCTGTTTCTATGACAAGTCAGAGGGCATCT
TTTTGAAGACATTCTAAACAGCAATTAAACTCAAAACATGTAATGACAAT
GACACACAAAACTCAACTGATGACCAAATGAAGAGTTCCAGCCAAGTTGA
CACAAGCTGGCTGACAGAGCTTGTAATACACACAGCTTGGCATATGCCTC
GCCATTTCAGAGATGTAAAAATAGGAATAAATGTTTTCCCTTAAATCAAT
GAAATAGAGCATTTGGACTGAAAATCTACGACAGTTATAGTGTTTTCTAT
TCATTATTCTCATTCTGTTTCTTCTCCCCCTTGCTTTCTTTTAGTTTGAA
TATTTTCTATCATTTCATTTTTCTTCCTACTAGTTTGAAACTTATGCATT
TATTTTCTATTTTTTAGCACTTACCTAAAATTACTCTGTAATCCATGGAT
CCTTAATTTATTTAAAAAACTAATGTTAATGAGTAGCTTTATTTTCCTCC
CATCTAATTTAAGGCCCACAGAACACCTTCACTTACCTCAATCCTCTCCC
AACTTACATGCTTTTAATGTCATATATGTTAATACCGTATACTTTTAAAA
CTTTCTAAAATAGCATTATTTTATAGCATGAGTGTTCATTTACATTTTTG
CATATATTTAGAATTTTCTTTGCTCTTCGTTTCTTCTTCTATTTATGACT
CCCCTCTGGGATCATTTTCCTTCTACTTGAAGTACATAGTTTAGAACTGC
ACTATTCAATACAGTAGCCACTAGCCATGTGTAGCTATTGAAGTTTAAAC
TAAGTAAAATTGAGTAATATTAAAAACTCAGTTCCTTCATCTCACTAGCC
ACATTTCAAGTGCTCAGCAGCCACATGTGACTAATGACTACTGTACAGCA
AACATATAGAACATTTCCATCATGGCAAAGAGCTCTATTGATAGTGTTCA
TCCAGAGTTTCTGTTCCAGGACCAAACTGAGGGTTGGGCTGCTATTTCTC
ATGGCCCAATAACAAGATGCAGATGAGCTGGGGAGGAAGAGAGTTTTTAT
TTCTGCAACCAGTTACAGGGAGAAGGCCTGGAAATCATCACCAGGCCAAC
TCAAAATTATGACGTTTTCCAGAGCTTATATACCTTCTAAGCTATATGTC

FIG. 4 (35 of 61)

TACGTGTAAGTGTGCATTCACCTGAAGACGTAAGTGATTAACTTCTTTTA
ATCTGTAACTAAGGTCTGAGTCCGGAAGATCTTCCCCTGGAGCCTCAGTA
AATTTACTTAATCTAAATGGGTCCAGGTGCTGGGGTAATTACCCTTATCT
TGTCCCCTGCTAAATCATGGAGGTTTGGGGAATTCCCTTTAGAGCACCAT
TAACCTGTTTGTTGAAGGCCTGGGAATTTCCTCCAAACCCCCATTAAACC
TGTTTAATCCCAAATTGGTTCCGTTAAAAATTCCCTCCTTAATTTGTCCA
ATTTTAAAGGCCCAAAAAAGGCTGGGGCAAACTCCTGAATGGCCTTTGTT
ACATTCCAACCTTTGTTTAAAAACACCGGTTTTTAATATTTAACTTAACC
ATTTAATCTCTACTGAAACACTTGTTATATAAATCTGCATTAATGAGAAC
TGGCCTGCGCCATATCTCCTTCTCAGAATATCTTAGGGTTGTGATCCCCT
GTGTGAAGAGAATATATCTCTGGAGATCTCAATCTCTCTACCCCAAAAAA
AATCTCACTCGGAGAAAACTCAGACTCTTATCTCCACAGCGCTATCTCTC
TCCTCTCC

>Contig50

GCTTGTCTAAGATGGTGCTCCTTGTTGCTGTGCCTGCTTTCATCCTGGGA
TCTCCCTTCACCATCAGGATTGCCTTCACCTCATTCCAGTCTTGGATCTT
TCTTCTTGTTTCTTGAGTATTTTTTTTTTTTTTTTTGCTGCATTCCCTTCA
GTGGCCTCTTGGGAAAAGATGTGTAGGGAGAAAAATTTTCTTTAGAAACT
TGCATATCTGACAATATATTTATCCTATCCTGACATTTGGTAGATAGTTC
AGCTGGGTACAGAATTCTAATTAATTTTCCTTCCTGATTTATAAGACATT
GCTCCATTTTCTTCTGGCTTCCAATATTGCTGCTGAGAAGTCTGACACCA
TTCAAATGCCTGATTTTTTCCATGTGATTGTTGTTTTCTGTCTGGAGTGT
TGTAGGATTGCCTCTTTATCTACAGTGTTCTGAAATTTCATGACGTAGGT
CTTTCTTCATTCATTATGGTAGACACTCAGTGGGCCATTTAATCGGGAAA
AACATGTGTTCTTCAAGTTCTACAAACTTTATTACTTCCTTTTTCTTGTG
TCTTTCTCTGGTCTGTTTTCAGCCCCGAGTCTCTTAGATCTGTCCTCTAA
TATTCCTATTGACTTTACTTCATTTTCTAAGTCTTTATCCTTTTGCTTTA
CTTTCCGAGAGACCTGCTTAACCTTATCTCCCAACTCTTTTATTGAATTT
CATTTCTTTTACTATATATTTTTTACTTTGAATACACCTCTCTCTTCCTC
ACATTTTCCCCCATAGTATTTTGTCTTCAATTGACAGTTCTACTATCTTA
TTACTCTGGAGATATTAATAATAGTTTTTAAATTTTTATTTATTTTTATT
TTCAAAACAGTGTCTTACTCTGTCACTCAGGCTGGAGTGCAGTGGTGTGA
TCATGGATCACTGCAGCCTTGATCTCTGAGCTCAAGCTATCCTCCTGCTT
CAGCCTCCCAAGTAGCTGGAACCACAGGCATGTGTCACCATACCCAGCTA
ATTTTTTTGTTTTTGAGGTGGAGTCTCACTCTGTAGCCCGGTCTGGAGTG
CAGTGGTGCAATCTGGGCTCACAGCAACCTCTGCCTCCTGGGTCCTGGTT
CAAGCAATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGAAACA
CACTACCATGCCCAGCTAATTTTTGTATTTTTGTAGAGACAGGTTTCACC
ATGTTGGGCAGCCTGGGTCTGAACTCCTGACTTGTGATCTGCCCACTTGG
GCTCCCCAAAGTGTTGGGATTACAGGCGTGAGCCACTGCACCCGGCCACT
AATTTTTAAATTGTTAATAAAGACGAGGTCTTGCTATGTTGCCCAGTATG
GTCTTGAACTCGTGGGCTTAAGTAATCTTCTGCCTCAGCCTCCCAAAGTG
TTGGGATTACAGGTGTGAGCCACTGAATCTGACATTTTTTAAAAGTTTTC
TTCTCTTTACCAAGTCTTTTTTCCCCTTTCTGCTTTTTTGGGTTGTTTTA
TTTTGATCTCTATCTTGCTAGAAACTTTCTGGAGACGTTTAGTAATACTA
GATTTTTGAGAGTGGGCAACTGGAAAGCTGATTGGAAACTCTGAATACAT
GGGTGAGGCTTGTTGGCTGTGAGTGTCATTGCTTGATGTCCTGGCAAGGC
CAATGGGTTTGGGACCCCTACTATTAGTATAGGCCTGATTCCCTGGGAAA
GGCTCTTTTGATCTCCTGCCTGGAGGATAAAGGCCTGGCTACCAGCCTTC
TGTGTGTAATGTGAGGGAGAAGGGCTGGAGTATTCAACATCATGCTGAAT
CCTTTCAATGATCATCTTGTTTTTAGTAATCTCCTACCTTAACTCTCTGT
CTTCTGCTAGTATGGGAAAGATGACCTGAAAATCTAACCATTTATTTTTC
CCCCATTAATATCATTTTATGATTATTCAGAAGTTAAATAATTGTCATGC
TGTCCTCCAAAAAGACTGAATCAACTAGCAACAAATAAGAATTTTCTCAC
AGCTCTGCCAGCATTTTAAAAGAATAGCTTTATTGAGCCCAGGAGGTCAA
GGCTGCAGTGAGCTGTGATTACACCACTCTACCCCAGCCTGGGTGACAGA
GCAAAACCCTGTCTCAAAAAAGAAATTTAAGGAACAGCTTTATTGTTGTA
AAATAGACATACAATAAACAGAGCACATATTTAAATTGTGCAACTTATAC
TTTGATATAACCCTGTGAAAACATCACCACAATCAAGATAGTGAATATAT
TTATCACCTCCTGATACAGTTTAGCTCTGTGTCCCCACCTAAGTCTCATG

FIG. 4 (36 of 61)

TTGAATTGTAATCCCCAATGCTGGGGGAGGGGCTTTGTGGGAGGTGATTG
AATTGTGGGGGTGCACTTCCCCCTTGCTGTTCTTGAGATAGTGAATGAGC
TCTCATGAGCTCCCCTTCACTCACTCTCTTTCCTGCTGCCATGTGAGGAT
GTGCTTGCCTCTTCTTTGCCCTTCTGCCATGATGTGTTTCCTGAGTCCTC
CCTAACCATGCCTCCTGTACAGCTTGCAGAACTGTGAGTCAGTTAAATCT
CTTTTCTTCATAAATTACCCAGTCTCAGGTGGCTCTTTATAGCAGTGTGA
AAAGGAACTAATATACCTCCTAAGTTACCTCAAGCTTCTTCTTAATTCCT
TCTCCTCCCTTCCTTCATTGCCAAGCAAACAACCACCTGTTTTCTGTCAC
TATAGATTAGTTTACATTTTGTGGGTTTTTTTTTTTTTTGAGACAAGGTC
TCACTCTGTTGCCCAGGATGGAGTGCAGTGGTGCGATCATAGCTCATTGC
AGCCTTGAACTCCTAGTTTCAAGTGGTCCTCCCACTTCAGCCTCCTGAGT
ACCTGGGACTACAGGGGTACACCACCACAACTGGCTTAAAAAATTTTTTA
AATAAAAATGGGGTCTTGTTATGTTTCTCAGGCTGGTCTCGAACTCCTCG
CCTCAAGCAGCCCTCCCTCCTTGGCCTCCCAAATTGTTGGGATTACAGGC
ATGAGTCATGACTCCTGGCCTAGTTTACATTTTCTAGAGTTTTGTATAAA
TGGAAACATACAGAATGTATTTTTTTTGCGGAGTGGGGGAGTGTTTCTATT
TCTTTCTTTCTTTTTTCTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCG
CTCTGTCTGTTGCCCAGGCTGGAGTGCAGTGGTGCGATCTCGGCTCACCG
CAAGCTCCACCTCCCGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCTGAG
TAGCTGGGACTACAGGCGCCCGCCACCACACCTGGCTAATTTTTTTTGTA
TTTTTGGTAGAGACGGGGTTTCACCATGTTAGCCAGGATGGTCTCGATCT
CCTGACCTCGTGATCTGCCCGCTTCGGCCTCCCTAAGTGCTGGGATTACA
GGCGTGAGCCACCGTGCCCGGCCCAAGTGTTTCTATTTCTTAACCAGCTT
TCATGCAATCTTTTTTTATTTTACCATCTCTGTGATCCCACTCCCAAAGG
TACTAGATGTCGATTGGTCCTTAGGATCAGCTACCATTTGCCCAACTGCT
TTCCAGCCTTCCAAAAATTTTTTTCTTTTTTTCTTAAAGATACTCCTGTG
TGAGGCTCAGAACTCTTGAATTGCTACTGCAAATATGAACTCGGTGATGT
GAATGCCAGGGAATTGCCTGATTGATCAAAGAAATGTATCCCCTTCTCCC
TCACTCTTGCTGTCTTCTCATTTGTTTTCCCCATCCTTGTGGATTCGTGA
ATTTAAATATCCCTTTAATGTTATAATATTTTAATGGCGTTTGGCGAAAA
GTACAGAATTAGGTGCAAGAGTGCATAGCTGTTATTTTTTTTTTGGCCTC
TGAGACTGTTCATATATGCAAGTTATTTAACAGAAAGTTCTGCAGTGACC
TGAGATGTCAGGGGGGTCTGATAGAGTACGTTTGAAGGCAGTTACTGGAA
AAAAATAATGCCATTTCTGGTTTGTACTTCGGTAAGTTCAGATGACCCAA
TATATTGTTTACATGTGGCATTCAGTAAAAAAGTAGCTTCCCCTCCCTTT
CTTCTTCCTTTTCTCCTTTCCTGCTTCTATAAAGCATCTGCTTTGGGAAA
CTTCTTAGGAGGAGAGCTTGCCAGCCCGTGGGTAATGGAGAGGTCTTGCA
GAGATAAAAGAGATGCTCCCACTCAATGCAGGATGGTGTGGAGGTAAATG
GGGATACGTCTGGCATCACTCAGGAATGGGCCTTCCTGGCAGGGAAGAGA
AGGGAGGGGAAAGAGGAAGGGAGTCAAAGATGAATTGCTGAATACGGGGA
TTCCAGGGCCTGGAGCCAGGAAGAGAACTTTGGGAGGTGTGAACCTGGAG
GGCATCAGCTGATGAGGAGCAGCCTGAAGTCCGGGGAGGACCTGTTTTTG
GTGGCCAGGAAGAAAGTGCCTTCCACACACAGGGAGGCCACAAGGCTGAT
GGGCTGGGGGTTGGAAGGACAGCCCTAGGACAGGCTTGGGAAGCAGGCTC
AGGTAGGGACTGCGAGGTTCTTGTTGAGTCTTTTTCATTCCTGGTCTTAG
AAAATAGAATCCAAGGCCTCTTGAGAGTGGAAGGTGGGTTGGGAGGAGGG
CAGATGGGGCTTAGGCCCAGGACACCCGTAGAGCTACTGCCCAGCTGTCT
CTCAGGGACTCTGCTGAGGTCACTCCAAGGATCATTCTTAGCCTTGCTAG
ACAGTACTGACAGAGGGAACCGTAGTATCGCACCCACTTCCTTCTCTTTC
AATGAAAGTTTAAAGGTCACCATTTCCTCTGGCAAAGGAAGTTCCACAAA
TATTCCATTTCCGGTCTTAGAAACAGCAAGGTATCAAGCAATTGCAAACT
TCCTGTGCTGGGGAATTCCCAAGGAAGTAGGGGCAGAGTTCTGGTGGAGA
CAAAGTGAATTCCGAGTGATTAGTCAGTAGCAGTAGCAGTAGCAGTAGCA
GTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGTAGCAGC
AGCAGAACCAGAATTTCCCCGCACGTGTCTCAGGCTCTCATTTGCCAACT
CAGTCTCTAAGTATTTTTATTGGCAGGAAAAATAAAATAGCTATGAGTGA
AATAATTCATTAGACCTGAGCCTCCATCAATTTTGTGTTTAAAGGCCTGA
CTCTCTTTACCTTTCCCTGGGATGGAAGATGCAAATGTTCCTGATGTCAC
TGTCAAAAAAGAAGAACCAGTGGGTATATTGTATGCTTGAGTTCCAGCCA
TTTGTCACAATAGATAGAGATGACTGCCATGTGTGTAGACTTTCTATAGA

FIG. 4 (37 of 61)

CTGTGTGCTAAACCCGACCTGCCACTTCCAAGGAGTAGATGAGGAATGAC
CATGGTTCTGGGGAGCCCTACCCCAATTTGGGGCAGACATTCCAAAGCTC
ATTTTCTGTGGAGGGGGTTGATGGTTAAAGGACGGCCTGGGAGTAACTCG
TCTGTACTAGGGCCCAGGAGAGTTACATGCTGCTTCCCATGTTATTCATC
ATTCCCCCATGTGAATAGCTATGGCGTGAGGTCCAAGGTTAGGGCCTTTC
TACCATAAATGGGGGAATAAAATTCCCCTACCAGCCTGAGAAGTTTCTGT
TATAAAGAGGCTTTTTTTTTGCGGGGGTGGGGGAGCAAGCCGACTAATGT
GTTATTTCCATACGGTTTGTTTTAAAATGTAGATGTCATATGCAGGAGAG
GTGGTGTAGTGAGTCACAACGGGATTAGAAGGACCAGTCCGAAAAGCAGA
AGAGGGTCAAGTTCAGGGCACTGAGGACTACTGCATTCAGTGGCGTGAAA
GGCAGATGGCTGAACAGGAGGGGGACATTACATTGCTTGTTCTCCTTGAG
CCTCGATTTCCTCATCTAAAAAGAGGGTCATTTATTCACAGAACATTTAT
TAAACTTGTGCCAGGCACCGTGCCAGGAGCTGGACTAAAAATTAAATCCA
CCCCTGTGAGCTGCTCTGAAGGCTAAAATATGAAGTATGTAAAAGTAACC
AAGTGCTGTACACATGCAGCTATTCAATGACTGTGTGGGCATTGCGGCAG
ATTTTAATTTTCTTTTTTATTTCTTTCTCTTTAGTGAGAGGTGTTGGTTG
TTATTATTGTCGTCGCTGTAACTGTCTATTTCACTTGCTTTTTTGTTGCC
TCCAGCCCATTCCAGGGCTGTCATCTAAGACACTTCTTATCACCTAAATA
ACCGGGGAGGCAAAGCGCTTTCTTAAGAGATGGATCCAGAAGAACAATGC
TGGTTTTCTGTAGAAAAAGGGGCTGTGGGAAGTAGAGATAAGAAGGGAAT
TGGCCAAGATGAATGTACAGAGCCTTATTTTTTTTTTATAACACAGCAAG
ATTAGATACAAAACAGGACAATAGCATCATCTGTTTTTATAACTGGAAAG
GACCTCACTTTACAGGTGGGGAAGAATAGAGTGGAGAAGTGAAGAGAATG
GTCACAGAGTCAATCAGCATGTCTGCGTCAAAGCTGGGATTCCCAATTCA
GGGCTCTTACTACAGTGACGTATGGCTAATATTTTGGCATTGTTTCGGGG
AAAAGCTGAAGCCCTGATGGTGTACGTCACTCTTGAGATAGTCTGTAGTC
CAGCAGGGAGGAAAGCAAGGAAGGGAGGTGGAGGCAGCATTTTTGGGTGT
AACATTTCGTTCTTGTTTTGTGGCCAAATCATAGTGTGATTGGGACAAGC
CACTGCCTTTCTCTGAGCCTCCACTTTCTTTTTCTTCTTAAGAGGGAGGG
AATAGTAGAGTAAAAGTAGTCATTTTATCAAACACCTGCTATTTTGGAGC
CATATTGCAAGTGGGTTGGGGGTTGAACACTTGGCTTTATTACCCATAGG
ATTAAATCCAACCTCGATACTGTGGCATTCCCAAACTCCAGTCTAATCTT
CTTCTCCATCAGCCATGCCCCACGACACCCTGGTCATATCTGATGTTGCC
CCTTGCACTTGCCCCCTCCTTATCTTTGCTTTCTGACCTACCATATGGCT
ATTGGTTGAAATTCTCATTTTCCAGGGCCTTGCTTAAATATCATCTCATC
CATTAAAACTTTCTTGAACCTCCCCTTGCCCTGTTCCTCCCTAATGTCTC
AAGCCAGAATTTATTTCCTTTTGTGGCCAAGGGACTGGGTTTGTGACCTC
TCTCACGAGACTTAATATTGAGACCAAACGTCTTTAGACCTCACCAGCCA
GAGAGATGAGCATCTATGGAATGCAGGCTTTTGCCTGGACTTGCTGATGC
AGGGCCTCTGCCTTCCTCCAGGGCCTCTCCTGCTGTTTTAGGAATTTCCC
TCATGGCACAGTCCATGAGCTCAGGGTCAAGTTCATACATGTTTTTACTT
CTTCTACTCTGCAAATGGTCTTCTTGAACTCTGAGGGTCCTAAAGCTGCT
CTGCAGTTTGTGGGGTGAGTAGAAAGGGGCTTTCAAAAGTTGTGCTGTTG
TTTCCCACCCCAATAGCATGAAACACAAAGATGCTTACAAATAGCTGCCT
TGCTTTCTAGTCCCAACTTCTCTCTCCTGAGGCTTTAAAACAAGTCCCCT
AGGTTGAGCTGGACTGGAGTTGTATCCTATCTTCATTATCTGTCTACTCT
CTTTCTGCTCTCTAGAGAAGATATTATATATGTGTGTATGTATGTGTAAA
TATATAATATCCATATATAGAACATATATTGTTATATTTACATATACATA
CATAACATATGCATGTATTCATATATACATATGTAGTATCAAAGTTGGAA
TTAAACTGTATATTTTGTAATTTGCTTTTATTTGCATCTATCACTGTAAA
ATGAATATTTATCCATACCGTAAGATATTCTTCAATGTATTTTTTTTTTT
TTTGAAACAGGGTCTTGCTTTGTTGCCCAGGCTGGAGTGCAATGACCCGA
TCTTGGGTCACTGCAGCCTTGACCTCCCCGGCTCAAGTGATCTTCCCACC
TTAGCCCTCTGAGTAGCTGGGACTAAAGGTGTGTGCCTCCACACCCAGCT
TTTTAATTTTTTTTGTATTTTTTTTTTAAAGACAGGGTTTTGCCACATTG
CCCAAGCTGGTCTTGAGCTCCTGGGTCCAAGCAATCCTCCCACTTTGGCC
TCCCAAAGTGCTAAGATTACAAGCATGAGCCACCACACCTGGCCTCAATG
TAATTTTTAATGGCTGTATAGTATTCCATCATGTGGTTGTACCCAAAATT
ATTTAACCAGTCCCCAGTTTATTTCAATTTTTTTTTACTATTTTGAATAA
TGTTTTAGTAAATACCCACAAAATATGTACAATGGCTGGGCTTAGTGGCT

FIG. 4 (38 of 61)

CACCCCTGTAATCCCAACACTTTGGGAGTCTGAGGCAGGTGGGTCACCTG
AGGTCAGGAGTTCGAGACCATCTTGGTTAACATGGTGAAACCCCGTCTCT
ACCAAAAATACAAAAATTAGCCGGGTGTGGTGGCACACACCTGTAATCGC
AGCTACTTGGGAGGCTGAAGTAGGAAAATCACTTGAACCTAGGAGGCGGA
GGTTGCAGTGAGCCGAGATCACACTACTGTACTCCAGCATGGGCAACAGT
GAGACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAAGTACAATTTGTTG
TACCTCCCTGATTATTTCTTTTAAGTAGAATTTTCTTATAATTTTTTTA
TAAGTAAAATTTTGAATCAAGGGAGAAGCACCTGGAGTCCTTCAGATACC
TATTGCCAAACTGAACTTTTCTGTTCCAGGTTTACTACATTCAGCCTGAC
TCAGGGTTTGGGGAGTAGAGGAGGGGGTGGAGGCAGAGGGCCTCTCCCTG
TCCCCACAGACCTCCCTTGGTGAGGTCCAAGTCTGGACAGGTGGAGTGTG
GCATTGCACCGTCAGGTCCTGCTTCCTGTAATTCCCCTAAATCCATCCAG
TGGAGCCTCATTGTTCAAGTCTTTTTTTTTTTTTTTTTTTTTAACTCCC
CTGAAGACGGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGA
TCTTGACTCATTTCAACCTCTGCCTCCCAGGTTCAAGTAATTCTCCTGCC
TCAGCCTCCTGAGTAGCTGGCACTACAGGCGTGTACCATCACGCCCGGCT
AATTTTTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAG
GCTGGTCTCGAACTCCTAACCTTGTGATCTACCCGCCTCTGCCTCCCAAA
GTGCTGGGCTTACAGGTGTGAGCCACCAGGCCTGGCCTCAAGTCTATTTT
TTAACTCCAGGAGGCCTGGTATTCAGAGGGATTAGGGCTGGCAGAAGGGC
CTCAAAGCTTTCAAGGCCTGGGGAATAGGCTGCAGCCTGGTTCAGGGTAA
CCCAAGTGATTTTGGTTCCAAAGGGACAGGAAAAAAAGTGATTGATATGG
AAGTTGTCAAAGTGCAACTGTCAAGACATTAAAAAATGTAACCCTTTTAC
TAATATACAGTAGACTTGTGTTAAATATTTAACTGATTGTAAAAGGAAAA
AACCAGACGCAGTTTTCCCTACCATACTGTCACAACACCTCAACACTGAG
TTCTTCTGTGACCTCTAGTCACCGAAATGCTTGGGGATTTCTCCCACCAC
TAGTCCTCCAGCAGCCGACACCAGTTGGGTGTCCTAATTCACTCCAACAC
TATCTACCTGGAGTTAGCGTTAGATCCCACAGGTTGAGGGCTCAGTCTCA
CAAGACTGCCTCCCACTTCAGGTGCCAGTTACAAGTGGTAGGTTGTCACC
TATGCTTCTGACTGATGGCTATAAATCTGGGTTTGCTTCCCTCGGGTTCC
GTGAATTTGCTAGAGCAGCTCACAGAACTCAGGAAAACACTTAAGTTTAC
CAGTTTATTCTAAAAGATATTACAAAGGATACAGATGAACACCAGATGAA
GAGATGCGCAGAGCAAAGCATGTGAGAAGGGGTGTGGAGCTTCCATGCCC
CTCTGGGGCACCACCCTCCAGGAACCTTCATGTGTCCAGCTATCTGGGAG
CCCTTCCAAACCCTGTCCTTTTTGGGTTTTTAAGAGTGGCTTTATTACAT
ACACATGATTGACCGAACCATTGGCCATTGGTGACTGACACAACCTTCAG
CCCCTCCACTCCCTCCAGTGGTTGGGGAGTGGGGCTAACAGTCTCAAGTC
TCCAATCCTGCCTTGGTCTTTCCTGTGACAAACCCCATCATGAAGCTACT
GCATTGGGGCTGCCAGCCAGCAGTCATCTATTAGCATGCAAAAGACACTC
TTATTATTCCAGAGATTCCAAGGGTTTTTAAAAGCTGTATGTCAGGAAAC
AGGAGATGAAGAACAAATATATATTTCACAACATCACACTCGTTGGGGGA
ATTGACAGGATAGCAAAACTGATTAAAGGAGGATAGGAGAGACTGAGATA
TATATTTCCATATATATATATAGAGAGAGAGAGAGATATTTCCATATATA
TATATAGATCTAGAGAGAGAGAGAGAGATAGAGAGAGAAGAGTCTTTCC
>Contig51
ACACATTTGGGGGAGCAGTTCCGGAGGTACAGCCCGGACAGGAGATGTGA
GAAGATCGTGGTTANTGTTCCCCTGGTCCAGAACCCCTCCAAGTGGGCTT
AAGTAGGAAGGGTGGTGAGCGGCAGGTAAACACACGTCAAAGGCAGTCTT
CCTCTCTGAGGGAAAACACTTGTATAAGCATTGCAATCAATGGGCCTCTT
TAATTATGTGCCAGTGGCAAGAGCGGGTGCTGAACCCAGGGGCCTGCCTC
AATCCGGGGCCTTTGAGGCAGAATAAAGTGGTCTCAGGTTGTTGGCATTT
CCTTGCCCTTCCACCCGAAGCAGACACAAATCCTCTCTGGAGGCAAGTTC
CCCAATTCAGCCAGTACAACTCCCACAGACTAAGATCAATCATGTACAAG
CTCACAGACAAAGGTCACCAAACACACAGAGCAATAAACAAATTCATGAG
TGACGTGAATGAGAATAAACAGAAACAATAACCACCAGCTGGGATGCTCT
AAGTCTTCAGCTGTTAGAATTCCTGAATATAGAATAAAACTGCCACAATG
GCAAACATGCATCTAGTACTTACTGTGTGCTGGGTTCTAAGAATTTTGCA
CATTGTGCCAGATACCGACTCAGCTTCACACTCACCCTCCTACTGTGCCC
TCTTAATTTGCACTAGATTAAAAGGTAGAAAGGAAGAGGCAGCTATTCTG
TTCTTGGCTGTGCCTCTGGCAGCACATGCAAAATGGGCAGTAACAGTGGC

FIG. 4 (39 of 61)

AGTCACAGGTAAGTAG[illegible] TTCTCACAG[illegible]TGGAGTTAAAGGCATGGGA[illegible]
GAGACGAGCAAGGTTCCTAAAGGGACAGTGGCCAGTAAATGACCAGGGGC
TACTGGAGTGGCTGCATGGCTCTGTGGAAGCTCAGAGGAGCCTTGGGTCC
TGCAGGTGCAGTAGCAGCTTTCTGTAGTTCCTGATCTCTGGGTCCCACAA
TCTTCCCCGTTTTTGCTCCTCCACTTCTAATTTTGTAACTGACTTCCCTG
TGTGTACTTCTCTCTCTGATTGAAATAGCCAGACTGGTTTCTGTTTCCTG
ATAAGACATTGTCTGGTACGAACACAGTAACTCATTTAATCCGATATCTC
TATGAAGGAGGTACAATAATTATTCCTATTTTACAGATGAGGAAACACAG
CAGAAAAATAAAGTCAATTGTCTAAGGTTGCACATTTAGTCAAGGGAAGG
GTTGATATAACATATAATTATTTAGAAAACATCTAAGGAAATAAAAGGCA
TAATTTAAAAATAAAACTAGGCAGGTTTAAAAAAATGAAGTAATCTATAA
GTAAAAAAGTATAATTGTTGAAATACATATCTTAGTGGATGGGTTAAATA
GCTGAAGAAATGATTAATGAACTGGAAGGTAGTTCTGAGGAAATCAGAAT
TCAGCATAGATAGAAAAAAATGGGAATTTACAAAAGTACACAGGAATTATA
AAAGAGGTTAAATTATAGGGAGGGTAGAATGAGAATTAACATTGGTCTAA
CTGGAATTTTGGAAGAAGAGAATAGAGAGAATGAACAAGGCAATATTTAA
AGAGGTGGCTGAGAATTTTTCAGAACCAACACAAACTATGACTTTACCAG
TAGAGAAAACAATGTACACTGAGGAGGATAAATAAATATACTATGAACAA
ATTGTAATAATAATACTCAACAAAGACAAAGAGAAGATCTTAAAATCAGC
AAAAAAAGAAAGTCAGACTTAGAAAGAAATGACAATGGCAGACTACTCAA
CAACAACAATGGAAACCAAATTCAGTGAAACAGTATTTTCAAAATGCATA
TTTAATCTATCTTTGAAGAATAAGGGTGAAAAGGGTGAAAATTGCTGCCT
TATACAAAATATCAACATTAACAAAAAGTAATGAAGGTAATATAAAAATG
TTTTCAAATAAACAAAACTGAGAGAGTTTACCACCAACAAGCATTCATTA
AATGGACTTTTAAATGCAGTTTTTAGGAAGAAGGAAAACAATTCCTAAGG
AAGGTCTGAGATGCAAAAAGGAATTATGAACAAAGAAATTGTTAAAATTA
TAGGTGAATTAAAAAAACTGCCTGCATAAATGATAATAATGACAATGATG
CTATTAATAATGAGTTGATAAGGATAAAGAAAAGGACAGAATTAAAATAC
TAGAAAACAAGCATGCTGGAAAGGATTCAGGAATTACTTGAAGGTTAAAG
TTCTAGGGTCCTTCTATCCTTCTAGAGGGGAGTCAATATATTAATTTTTG
ACCGTCACTTACACAGTGAAAAACTTTAAGGATAACCATAAAAAAATAGA
AATAGAGAGTATAACTTCTGAAACAGTCAAGGGAAAAATATGGAATAAGA
AAACTGACCAAAAAACATCTCAGTCAATCAAAAAAAAAAAAAAAAAAGAAA
GAAAAGGTTCGGAAGGAGAAAATCAAAGCATAGAAAAAGCGGGACAAATA
GAAGTGGAAAAGAAAAAGGTAGAAGAAACAGGTCCAGAAATATCACTGAT
GCACTAAATCACCATTAAAAGATGAAAACAAATGAACAACATCAAAAAAT
TCTAGTGACTGTAGTAGTGCTGATCAGAATAGGCTCTAAGATAAGATGCA
TTATTGTGAGTCAACTTGTGATGATGAAAGGTTTAATTCACCAGAAAGAC
ACAATTATAAACTTGTAATCAAATAGTTTTATTTTATTTACTTTATTTAT
TTATTTTTTTGAGACAGGATCTTGTTCTGTTGCTCAGGCTGGAGTGCAG
TGGCTTGATCTCAGCTCACTGCAGCCTCCACCTCTTGAGGCTCAAGCTTT
CTTCCTGCCTTAGCCTCATGAGTAGCTGGGTCCACAGGCACACACCACCA
AGCCCTGCTAATTTTTGTATTTTTTGTAGAGATGGGGTTTCACCATGTTA
CCAGGCTGGTCTCAAACTCCTGGGCTCAAGCGATCTGCCCCCCTCGGCTT
CCCAAAGTGTTGGGATTATAGGCGTGAGCCACGGTGCCTGGCCTCAAATA
ACTATTTAAGTGAAACAAAACTAGTATGGCACTAATGAAAAATGTATAAA
TCCATAATCGCAGAGGGATTTCAACTTACTTCTTTCGATTATGTAAAGGT
CAAACAGACAAAAGACAATGACAAAACTTAATGCAATGAACACTTTTGAT
TTAATGAACATATATTGGATATGTACCCAAGAATTAGAGAATACATACTA
GTTTTGAGTTTATGCAGAACATTTACAAAAATTTAGTGGAAGCCTAAATT
ATAAAAAGTTGCTGTCACGTAGAATAACACACAAACCCCTGAGTCCGGAA
TTCAAAGCCCTCCACACTCTCCTCTACCTTTGCATCTTTATCCTCCACCA
CACTGCAGTGCATACTCTGGGCTACTACTCACTGTTCTTGATTCAAATTC
CATGTTCTGTCAGCTCAAATCATTCTCTCTGCCTGGAATAACTACTTCAT
ACATATTCTGCTATTGAATTCTTGTCTTAGCACCCCATCTACTCCAAGAC
GATGTCCAGTTGGGGTTACTCCCTGTCCCATTTTCTTTGATTACACTTTT
TTTTTCTACTTCCATTATATTATTGATCACATCTGTGCCACAGTTTTTGA
CTTTGTGTCTGCTTTTACTCTTTTCTAGACCCTGAGAGCTCCTGAAGGGT
TGGGTCATTTCTTTTTTATTTGCTCATTCCTCATGGCACAGTGAGTGCTT
AATAAATGGCTATTGACTGAAATTAAACTGTATCTAAATGGACATATTCC

FIG. 4 (40 of 61)

94/118

ACTTCTGGGCCATTCATTCTTTCTTTCTATTGGAACCAGGAGATGGGGAA
CCATAACAAAGGTAAGGTTGTGCCATGTGAAAGAACATGGAACCTTCCCC
TGAGGGCCAAAAAAGAGCAGGGAAAGGTGCAAAGACAAAATCTTCCATTT
TTAAACAATGTAAGAATGTGGTCCACCTCATGCTCAGGTGGGACTTTATC
ATGACGTTATTTTTGGGGACTTATAGCTGCATCATTTACCCCATATACAT
TTACCTTTAGTGTAGGGAACTGAGGACAGGAATTTTGTTGATGCAGACTC
TTGCTAATGAGGCTAACACTTGGAGAATTTTTATCATGCATTCAAGAAGC
TTGTTTTACATTTCTTCATTAATACTTTAGTTGGTGGTTTAGCTTTAGTT
GTAGGCTTATCAGATATTTGGAGATATCTTCATAAACGATGGCTTTGGTT
TTAGAAGAGTTATTCTGAAGCTACTATTTCTGGCAATAATCAAACAGCAT
GGCCATTTGTTTTGTAAGGCCTTTCCTAAAATATGACGGTAAAATCTACG
TGTGGAAAAAATGCTTATTCTTCTGTCCTCTATAAATGTGAATCTAGTTTG
TCTTCAAAATGAAATCAAGTGATTAAAATGTAGTTTTCTAAGAAGATAAA
TGGAGCAAAGCACTCTGTGTTTCACAGTGTTGGAAATCACTCATCCCTCA
TAAAACTGTCCCAACTGATCCTGACTCACATGAATGAATTAAAATAAGAG
TTAATAACATCAATTTACATTTTTAAAGACACTTTCCCATGTTTTAGACT
ATTGGTTGGAAAAGCTGGTAGGTGTACAATTTGTGGAGAGTTGGCTGTTT
TTGTCTGTCGTTGTTTGACGTATTTCAAAGCCATATCTAATTTTGTTGCA
GAATGGTCTGAATTCTACAAAAATGTTGAGTTGTGTAGTGTGGAGAAGTA
CGGAGCCATTTACTGAAAGGCTGGGGGAAATGACGAGACCCTGAGATAA
GGCAGTAGTGGTGCGAACAGAGTGGAAGGGAGGTAGTTGAGATATGTTCA
GAGTAGAATCAGAATGGACATAGTGAACAACTGGATGCAGGTGGGGGCTG
AGGAAGCAAAGTTGAGGATAATTCTGAGACTTCTAGGTTGATCCACTGAA
GTTACATTATTCAACACCACAAGGAAACTAGGGGAATGAGAAGGCATACT
GGTTTGCTTTGGAGTGGAAGGGCAGTGATGTAAGAGGAGTTAATGAGTTA
AAGTTTGGATATGCCTGAACTTCAATTTGATATGTGCATCTGATATACCC
TTGGGGTGACCCTCCAGGCAATGGTTGAACATGTGTATTTCTTAGTAACT
GATAGGCATCACAGACTCACATCAGTAAGGAAGCAACAGCAAACTTGATT
GGACGATATACCTGGAACTCAGTACCCTATGACTGGAGCAAGTCTCTGTC
AGTGAAATGAGGATAAGAAGAATCTTGACCTTGTGGAATATGTTGTTAGG
AATATATGTGATGAACAACATAGGATACTTCCTACAGGGCTCCACATGTA
GTAAGGGCTTTATAAATGCTTGATAAATATTATTGTTGTAATTTATTTCC
AAAGTAAGATGCCACTGGAGGAATCTTTGGAACCCAAATTAATAACAAAT
AGGACTGGATGCAATGGCTCACACCTGTAATCCCAGCACTTTGGAAGGCC
AAGGCAGGAGGATCTCTTGAGCCCAGAAATTCAAGACCAGCCTGGGTGAC
ACAGGGAGACCTTGTATCTATGAAGAATTAAAAAAAATTAACCAGATGTG
GTGGTGCACGCCTATAGTCCCTGCTGCTTGAGAGGCTGAGGTGGGAGGAT
TGCTTGAGCCCATGAGGTTGAGGCTGCAGTGAGCCATAATTGTGCCACCA
CACTCCAGACTGGGTGACAGAGTGAGACCCTATCTCAAATAAATAAATAA
ATAAATAAATAAATAAGTACAAACCAGCAAACACTAATCCTTTCTAGAGA
TTATTGAACTCTGGAGGGCAGATCTGAATGGAGCCAGCAGAGGGACCTAT
GGAGATCAGCCTGGCCCTGGACAGCACCAGGCAATGGGGTTGCTAGAGAG
GTAATGGGGTTGAACAGGGTTTAAGCCATGAGGTCTCAAGAATCCGTGAA
GACTCAGACTAATTTTTTTTTTTTTGCATGAGGATTAGGTGTTCCTAGGA
ATTTCAATGAGAGCAGGGTTAATGAAGGAATGCAGGGTAGGAGAGCTGAG
GGAAGGCATCTGAGAGAGCCTGGCTTATGAATGGCTGCGTCAGTATGGCT
CACCTGCTTTCCTTGTATCTACTTAGCAGATGATCCCACCCCAGGCCTCC
AGGGCCAAGGTCATTTCCACATAGTCATGGGCCCTTGAGGGCCTGGAGCA
GTGTAAGGAAGACAGAGTCTTAAGAAATTGCATTAACAGTCATGGTGCTT
GGCAAGTGTCGTCATCCTATGCCAAGCCTGATCTGAAGGGGTGCATGCTC
ATAGGTAGCTGCTGCCCAAGATTACAGCAGCTTCTTCAATCCCAGATCCA
TGCTCTCCTATATTCATTTTTCCAGGGGTTCCTGTCCTTCGACAGTGATG
AGATGCAGAATGACTTATTGAGTTATTCTCCTGATAGTTGCCAACTTTTC
CAAATGACAATGGGGCATGGAGCTTGAGAGTGGAAATGAGGCCCTAGGGA
TAGCGTGCTTAGGAAAACACTCCCAGCCTGATGTAATTCTGGGGGTACAA
TGGCATTTTCATCATCAAGACTGATGTAAAGGGTGACTAGCAGTGAGTTG
GGGGTGACTCGCACTGGGGCTAGGTTTCTGATTCTGCCTAATCCAGACAG
AGCAGAAGCACTAGTGGGCTGGTAGAGGGCCTCCAGGGCCTCACTTAATG
TCCTGGAAAAACAGCTCCAGATTGTTGGTTCACGTTCTGAGGACAAGCTT
GGGTACTACAGGATAGAGAGAGTGGTGGGAGATGCCGTGGCCTGCCCTGC

FIG. 4 (41 of 61)

TGATGCCTGCCCTGCCATTCCTGCGTGTGATGTCTCTGGGGCATCTTGCC
TTCCCTGCCCAGACCTGTAGTTCAGCTGAGGGCATGTGGAGGCCAAATGG
CTTCTTAGAGTGTTACTTTCCTTGAACAGCTCTGCTGGGAGAACTGGAGG
AGCTAGCTAGTCACGGTAACTGCAGCAGTCAAAGGATCGTCCCGGTGGAG
GTGGGGTGGAAAGGTAGAGAAAGAGAACATATAGCGTTTTCCTTGGAGAT
GTGTGGGCATGTCATAGAGGAAATACCCAATTCCTGAGCCTTGAGCCCTC
CAGGAAACCTTGGAATATTAGGTTAGTCATCCCCAAGGAAGTCTAAGAAT
TCTGGTCTCACCCATCTCCTTTAATTCCCACAATGATCCTACATGATATT
AAGGAACACGGGCCAGTAACCCTCCAAGCAATGGATGTGGTGGTGAAGTT
TGACCTCATGATGGAGCGGAGGTTGGTTTGAAACCTAAGAATTTAATTTA
TTGTTTCAAACTGTTCTCCACTCAGCGTTATTAAAGCATACATAATTGAC
ACATAAAAATTGTATATGTCTACGGTGTACAATGTGATGTTTCGATCTAT
GTATACATTGTGAAATGATTACAACAAGCTAAATAACATACCCATTCATC
GTGTTTCAAAGGAATTAAACTCAAGCACAAAAGAGAGGTGCTGTTGAAGA
GTAGGGCTGCTCTATCTAAGTAGTATGTCTGGGGTTGTCCTGGATCAGGG
TCCTTTTGTGCTAGTAATAAACCAGCCCTTCTGGGGCTGCTCCACTTTCC
CCACATTTTCTTCTGGAGCCTCCCTAAGAATTAGGACATGGCCACTTTCT
CTGCATAGGCTTCCTACTTCAACAAGGACAGGGCTTGTGCTGCCCCATGC
CACTTGAGTGTCCCTACAGCACAGAGCTGAGTGCACACTGGCTGAGTGAG
GAAATCCCCCAGATTAATCTTGGTTCTAAGCATCATGGCTGTATTTCACA
CGTATATGAATTACAAATTACAGCATAGTCGAATAAGGATTTTTGTGCTA
CAACTGGAATCCCAGATTATGCAAATTGGATAGTATAATATTGAAATTCC
TAGGACTTTTTATTAGTTTTAAAAAATTATACAAGCTTAGAGTAAGAAAT
TAAACAGTGCAAAAGAATTCACTGTGAAAAGTAAAATGCTCTGTCTCTGC
TGAGAGACAGATATTGCAGCCCAGATACTACTGGGGTCAATAGTTTCCTT
TAAGCATGCCATTTTGATGGTTTATGGGACTTACAGCTCAAGAAGCTTGA
CACTAGGGTTGATCTCAGAAAATCATTGTTGCAGGTATTAGATATGACCG
TCTCATAAAGATACACACACAGACACAGCGATTGGAGATATTCACTGGGG
CTTATGGGCTGCTTGTCCTTTCTGCTCTGTGCCTAAGTTGGGCTCAGAGT
AGCCTGGCATCGGCTGTGGGGAGAATGCTGGCATGGGGTTAGCAGGAGCC
CACTTAACATGTCCTAAGCCACCTGGAAGAGTCCTTCAAGGAGACCAGAC
TCCAGAGGCCCTAAGGAAGGAAGGACTTTTGCCCGTTTTTAGGTATTCTA
GTCCCAGAGTTTAGGGAGGAATGGTTTGGCTTTGGGTCGTGTGCCCCTTT
ACCGAGTGGGATGGGATGTGCCCATGAGCTGTTGAGCTGGCTCTTGGAGA
AGACAGCAAAAGCGGGAATAAGAGGTCAGGAAGCTGTGTGGTTGTAGGAA
ATCCCAGCAGAGGGCCTGGGGGTCAAAAGTGGTCATGGTAGTGACGGTGG
AGGCTGAGGTGGTAGAAAATCAGAGGACAAACCCCATGGGCTGCTGGTGA
TCTGACCGAGCTCCTATGCTCTCCTGGTTCATTTAGGCTCTGTAGCAGC
AGATGATTGGCTGGTGTGAGAGCAGTGCACCTGCCATATCAGGCAATCCA
AGACAAGTCCAAGCTACGCTGGGAGGAAACCTGAAGGCAGCAGCAGGTAG
ACTGGCTGAAGACAGACAGGCAGGCAACTTGTCAATCAGATTTGTGTTTT
TAAGGACTTTTAACTGGGGAGCCCTCCATGACAGATCAGATGAGAGAGGA
ATCTGGGTCCGCCCATGTGTCAAGCTACCAGAGGGTCCCATCGGTGCTTG
GATCTTCTTTGAAGCTGGGTCTGAGGTTTGCAGGTAGAGGGTGAGCTGGT
CAGAGGGACCTATTGCAGAGCTAACCAACACCTTCCCAGGAATGCAAGCA
CAAGCACCCCACCGCGGGCAGGCGGGCAGGCACTTCTCCTTTTGCCACCA
GGACCTCACAGAGGCTGATCTGGCTCTGTGAGGTGGGAAAATGGGTTGTA
CTTAGTACATAGAGATAAAAGGCTTAGGAGGCCCCTCCATCCTGTGACCC
TGTCCCCAGACCACAGGTGCCGGCAGGTGCTGCTATTTCAAGGCTGGGCC
TCAGTGCAAGCTTGTGGTTTCTTGCCCACCTGTGATGTCCTCCCACTAAT
GAAGGGGCTCTCCATCCTCTGTCTGCCTCTAGCAAGTGGAGGCTCTGGGC
CCTGGGCAAGACACAGGGGGAAATGCCATCTGTTATCCAAATATATTTCA
ATGTGACAGGAAAGCTGTCTTTAGAGCACAGC

>Contig52

GCATGTGCTCTACATTGATCCCAGGAGTTTGAGACAACATTGCAAGACTG
GGCAACAAGCAAGACTCTGTCTCTACAAAAAATAAAAAAAATTAGTTGGG
CATGGTGGTACATGCCTGTGGTCCCAGCTACTCCTAAGTTGAAGAGGGAG
AATTGCTTGAGGCCAGGAGTTCAAGGCTGCAGTGAGCTATGATCACACCA
CTGCACTCTANCCTGGGTGACAGAGCAAGACCCTGTCTCTAAAATAATAA
TCGTAATACATTTTTTTAAAGTAAAACAAAAAAAGGTCACACTTTCTCA

FIG. 4 (42 of 61)

TACCAAAATAAATTCCAAATAAATTAAAGGCTTAAACATGAGAAAGTTAA
ACCATAAAATTACTAGAAGAAAATAAAAGCAAATATTTAGATAATCCTGG
GGATAAATTTCTTTGGAATGAATTTCCTTAAGATGAATCTCTAAAAGTGA
AATTCAGGGTTCAAAGGTCTTTTCTTTGTCCTTTTCTTTTCCCTTTCCCT
CTCCCTTTTCTTTCTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT
TTCTTTCTTTATCTTTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTTCC
TGGTTGCTTGCTTTCTTTCTTTCCTTCCTTCCTTCTTTCTCTCCCTTTCT
TTCTCTTTCTTTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT
CTTTCTTTCTTTCTACTTTCTTTCTTCTCTTTCCTTCCTTCCATCTTTCT
TTCTTTCTTTCTTTCTTTTCTTTCTCTCTTTCTCTCTCTCTCTCTCTCTT
TCTTTTTTTTTCTGGTGAGACAGGGTCTCATTCTGTCACTCAGACTGGAG
AACAGTCGCATGAACATGGCTCACAGCAGCCTTGACCTCCTGGGTTCAAG
CAATTCTCCTGCCTCAGTCTCTCAAGTAGCTGAGACCACAGGCACCCACC
ACCAAACCTGGCTAATTTTTGTATTTTTAGTAGAGATGGGGTTTCACCAC
ATTGGCCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCTGCCTGCCTT
GGCCTTCTGAAGTGCTGGGATTACAGGCTGGGCCTCTACGCGCGGCCGAG
ACTACCTCTCTTTTAACTGGATCTCTGAGCTCTGGGCAGAGCCCACCCTG
AATCCTGGTCTCCAAAAAAGGGAAAATTATTAGGAGGCTAGACCATATGAT
GCTTTTACAGTGCACTTAAAAAAAAAGTTTGTTTTTTTTTTAAAAGACATT
TCTACATGTCTAAACTACAATCTTCCTTGAAAACCCAAGAGTAGCTTCTG
TTGCAATAGCTAGTCAAAAATATAATAGTCAAAAAAATCAGGTAACACAA
CACAAACGCAAGCAGTTTAAGAGCTGAAATGAACTTGTCTGTTTACACTC
TAGGGATTCCATAAGGAAAAATAGAAGTTTCTCCCTAAAAGGGAGCCTGG
CACCTTCTCCATTTTCTTTAAGGAACCCCAGGCTATTATAAACTATTTTA
GGGCTCTCATGCAGCAGACGGTGCAAGAGAAAGGAGAGACAGCAGAAGTA
AATGAAGAAAACAGAATCCAGTCAACAGAGAAGAAAAAAACTTTTGCTCA
AAAAAAGGCAAGTTCCTAGGAAAGAAAAAAAAAACATGAGGGCTATTTAA
ATACAAAGACGCATACATACACATGCACACATCTTGGATGTTAGCTTTTA
ATTAAGCTGACTTTTAACTATTGAGGTCCTTTAAAATAAATCTTTTAAAA
TCTTATTACGATATTTCAGCTAGGACAAATTGCTGCTATTTCAGCATTAC
CAAGTATCAAACCAGAAAAGGCTTGATTTAGGAACCAAACCCAGGCTGTC
GTGGTAGGAAAAAAGGCAGAACGTTAGCTATGGAACCCACAGCATGGGGC
AACAGCCATTGCTCTTTCAGTATGGCCTGGCTAGCAAAAAGGTGGCCTTG
TTATGTAAATAAAGCCCGTTTGGTGGTCAAAATGAAACATCTTTTCCTTT
TTTTTTTTCTTTTGCTGGCCGTTTTTTCCCCCACCATACCACGTTTGTGT
GTGTGGGAGGGTGGGAATTTAGCCACTTCAGAGGCCTCATTCCCCATAAT
TTGGAAATTTCCTTTGGATTTGATCAAGTCAGATAGAGTAGGTCAAACCC
AATGGGAAAAAGACTGAAACAGCAATAAAAACAGAAACAAACAGTTAAGC
AAAATGAATGATCACACAACTTATATGATTACTGAGTGCTCTAATGGTAA
GGAGAAATTAAGACCAGCTGGTTGTTAAACTTTAGCCAAGACAAAACCCC
AATTCAGCTACTTACCTAGGGTTGGGTCTCAGGCTGAAGACCGCTCACTA
CCGTTCTAGAAGCAAGAAATAAAACTTGAACTCGTCTTACCTGTGTAGCA
GGACAAGCCGCAGACAAAATCCCTCAGACACCAAATTAAAGAAGGAAGGG
CTTTATTGGGCCTGGAGCTGCGGCAAGACTCACGTCTCCAACAACCGAGC
TCCCCGAGTGTGCAATTCCTGTCCCTTTTAAGGGCTCACAACTCTAAGGC
GGTCCACATGAGAGAGTCGTGATAGATTGAGCAAGCAGGGGGTATGTGAC
TGGGGGCTGCATGCACCTGTAGTTAGAATGGAACAGAACATGACAGGGAT
CTTCACAGTGCTTTTCTTATGCAAATAACCGATTAGATCAGGGGTCGATC
TTTACCAGGCCCAGGGTGTGTCACCGGGCTGTCTGCTTGTGGATTTCATT
TCTGCCTTTTAGTTATTACTTCTTTCTTTGGAGGCAGAAATTGGGCATAA
GACAATATGAGGGGTGGTCTCCTCTCTTACCTGCGGGGAGTGAGCTCAAA
CTCCTTAAAGGAGTTACCTGCCTTCCATCATCAGGGAAGCAGGAAATCTT
GCCTTCCTTGTTGGAAGCAAGTAAAACTCAAAACAAACAAAGAAAAAAAC
AGGGAGTTGTACAGCAAAATAAACTTTTGATTTTGACCAAATTTTGGGAG
ATCAGGAATTCTCTGAAGGAGATGCTTTCAGACCTCAGCAAATTGTCCTG
TTGGTTTGAGCCATAAAGTTAGCTCATGCTGGTACCAAACACCAGTAGGA
GATTTGTCAAAGGTAAGAGGCATCTCCACTCAGAATCCCTTCGTGGTTAC
CAACATGTGAACCTTGGAAATCTGAGACAGGTCTCAGTTAATTTAGAAAG
TTTATTTTGCCACGGTTGAGGACACCCACCCATGACAGAGCATCAGGAGG
TCCTGACCACATGTGCTCAGGGTGGTCTGAGCACAGCTTGGTTTTACACA

FIG. 4 (43 of 61)

TTTTAGGGAGACATGAGACATCAGTGAATATATGTAAGATGTACACTGGT
TCCCTCCAGAAAGGCAGAACAACTTGAAGCAGGGAGGGAGCTTCCAGGTC
ACAGGTAGGTGAGAGACAAACAATTGCATTCTTCTGAGTGTCTGATTAGC
CTTTCCAAAGGAGGCAATCAGATATGCATTTATCACAGTGAGCAGAGGGG
TGACTTTGAATAGAATGGAGGCAGGTTTGCCCTAAGCAGTTCCCAGCTT
GACTTTTCCCTTTAGCTTAGTGATTTGGAGGCCCCAAGATTTATTTTCCT
TCTACATCACTGTGGGCAGCTGACTAGGAAAGCTTTGTAGGACTGGTGGG
CAGTGTGAGAGCCCAGTGGGGGGTGGTGGTCCTGTGCCAATGGTAGCAAC
CACCTGTGAGGCTGAGTAAACTCATTTCCCAACCTCCTCTAGCAGCCCCA
GTGGAGATACAGATGAAGCAGACTAGCGATACAACCCAGCCTGAAGTTTT
GTCTGGTGAGTGTAATGGAATAAAAATGGGAAGGGTGCTGAAGAGACCAG
CAAGAAAATGGTTGAAGAGATGGGCACAGAAATTAAGCTGGATCAAAAA
GGACGGAAAAGCAGAAAGGGCCGATAGAGAGAGGGGATATCTATGGGTTC
GCGATTCTGAAAAGGACAAATCACTGGTGCTTTGAGAAGAGAGAGGGTGA
GAAAGCAGGAAGGCTGGAGGCTGTCATCCAAGAGGCGGACATCTGTGAAC
ATGATTCCAAGAGTCACCAGACCATGGGGGTGGCCAAAGGGAGTGCCTCT
TCTCACCTCCTACTCTTAATTCCTTGTACTCAAGATAATAAGTTCCCAGA
AGAGAAGTACCCATATTTAATTCATCTGTGTCTTCCTAGCAGTACTAAAA
ATATTATATGAAAGGTATCAAACCTTTGAGAATGTGTGCTGCTAAATTGT
TAAGGATGCTGGAAAACTCAAGACGTCCCTGATCCTGAGCCTGAGTATGA
GCCTGTGGTGAGCCCAATGCAGGTCTCCATTCAGACAAAGGCCTCAGGGA
ACGGATGAGACCTAGGGACAGAGATGCATGCTGGAGCAGCATTCCCCATC
CCTACTGCAGCTCAGGCCAGCTGACTGCTTTATGAGTAAACGTTACCAGG
GAACACTTTGCAGTCTTAACACACATGCCCACCTGTGACCACTGATCCCT
GTTGGGTGACCACTGACATCAGAGATTCGATGGCAGCAATGAAGACAAGG
CTATCCTCATTAGGAAGGAAAGGAAGGAGGAGGGAGGAGGGCAAACGAAT
CTTTCCTGCTTGTCAACCACGTCCATCTCTGTTAGGTGATTTCCCATGTG
TGACTTTGTTTATCTTTATAATAACTCTGAGAGGTAGGTCTTGATGTCCA
CATTTTGAACATGAGGACATCCAGCCAGGAAGTTGAGTTCTGGGGACATA
GCTGAGAGGGCAAAGCTACATATAAACCCCTCTTTGTTTTTTCTGGCTTA
TCCACTGAGTGCCCCCTGCAATCCACCAGCCCATTTGTGAAGTGCATACT
ATAGGTAAGTTGGCACAGGAGGAGTGGATGTGGGCGATTTTGTCACAGCT
CTCCAGGAACTTACACACTGGTGAGGAGGGCCAGGTATGTTCCTGACCAG
TCACAATCAAAGCAACCTCCTACTAATCAGGGAGGCTTGGTACCTGGGGA
ATGCTATGTTGAAAGGTTCTTTTCTGGGTTTTAAAATGATGGGTCTATTT
CCTTATTCTTAAGATTGCTTTTTTTTCTGGCTAGAACTTAAAAGAAATTTT
CAGTAAAATTTCCCTTCCCTGGCACAAAGTGAGCTTGAAATGAATTCCCA
GGTGGCCTTGATACTTTAAAATATTGCCTCCTATAAAATCAACCTTTAGA
AGAAGGAAGTCAAAGAACATGCTAGATTTCACAAAGGTTAATTCCTTGAA
ATCCAGTTATCTACAGGACAATGTTGTCAAAGAAAAAATTATTTGGCCAG
GCACGGCGGCTCATGCCTATAATCCCAGCACTTTGGGAGGCTGAGGCAGG
TGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAA
ACCCCATCTCTACTAAAAATACAAAAAAAATTAGCCAGGTGTGGTGGTGG
GCACCTGTAATCCCAGCTACACGGGAGGCTGAGGCAGGAGAATCGCTTGA
ACCCGGGAGGAGGAAGTTGCAGTGAGCCAAGTTCAAGCCACTGCACCCCA
GCCTGGGCAACAGAGCAAGACTTTGTCTCCAAAAAAAAAAAAAATTCAAT
GATATTTTTAAATTCATGGTAAGGAAGATTTCATTCAGAACCAGCACAGA
AGATATAGGAAACACTGCAATGGGACTTTGCGGTGGGGGAGAGAGATTGA
ACACAACTACATATACAGCACGGGCAAGGACATATTCATAGCCAGGAAGC
AGAGCAAAGATCAGTGGATGCGAAATTACTAAGAGGAAACATGAAAAATA
AGGGAGCTTCTGCCTAAACCCACCTAACCGGATCCTTGCTGAAGACAGGA
CAGGGTGATTGGACACCACTTTGGGGATGGTGGAGGATGGGGAATCCAGT
GAGATTTCAAGGGTGATCAGATATTGAACATAGAAGGTTCTTGCTAAAAA
AGGAGTTTACAAGAAAGTGTACAAATGTGCCTGGGAGAAGGTTCAGGAGC
CTGACTAAAATTTGGTCAAGCAGAGAATATTTGCCAAGATAATAGCTAAG
TCTTCTGACAAACAATAGATGCTAAGCCAGCAAGGGTGATGTGCTCAGAG
AAAGCACTGAGGGCTTATTTCCTTTTCCCCCAATCTCCACTCAGTCAAGT
CTAGTCCCCTTGTCAATGTAGCCATTTGTAAGAATGCAATCAGGCAGGGT
CCCATCTCCTAGTGACAGGACTGACTGAAGTTCTGCTGAAGAGAGTGGCC
TGGGGCTGACACCGAGATTTCAGAGTCCTGGGTTTCGCCGAGAGCTCAGT

FIG. 4 (44 of 61)

GTAGTGCCATGCCCTCTCTCCACCTGAACGCCCAGTGTGGGCAGGAACAA
CTGCAGCTAGAAGTCTGGCACTTACGCTGGGGTCTAAGACCTGCCTGATC
TGCTAACTAGTCTTGTCCCTTGGCTATAAACTGACGTTGGCACCTGGCCA
GAAAGATGAGCAAGAGATCTCTGACACACCTTTAAGTCCCTGTGGAGTAG
GATTATGTTGGGGAAGGTCATTCTCTTGACTGAGCAGCAATTTCAGAAGG
AAGTCCCATGCCGAAGTGAGAGAAGGCAGGGAATCCTGCCTAGTCAGCTA
GAGCAAAACAGTCTGCAGGACGGGACCCAGGGATGTGATCCTCCCATCCA
AAGGCACTGAACTAAATGACTAAAATACTTTCCAGGGCTCACGTTCTTTG
AAGAATGGGGACTAAAACTAAGACAGGAGCCAGCAAGTGAGGACTTGGAA
GGAGATGGCTCATCTGATCAGCCTCCACTCAACAATTTTAATCATCCACA
CTGGCATGGGGACACAATATGAATAAGTTGACAGGGACCTACTCTGATTA
AGCAGTGGGCTAGTGCAGAGACCTGTCAGTCAAGAGTGGACAGGAGATGA
TTTCAGACAGTGAGAACAAAATTAACAGAGTCATGTGCTAAAGGGTGGCT
GGAACTACAGAGGAGTTTAAGACTCAAGAGGTCTGGCTGGGCGCGGTGGC
TCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACAA
GGTGAGGAGATCAAGACCATCCTGGCTAACGCAGTGAAACCGCATCTCTA
CTAAAAATACAAAATATTAGCCAGGCGTGGTGGCGGGCACCTGTAGTCCC
AGCTACTCGGGAGGCTGAGGCAAGAGAATGGCGTGAACCCGGGAGGCAGA
GCTTGCAGTGAGCCAAGATTGCGCCACTGCCCTCCAGCCTGGGCGACAGA
GCGAGACTCCGTCTCAAAAAAAAAAAAAAGACTTGAGGGAGTTGTTTATT
TTTGTTTTCTTTTTAAGACAGGGTCTTTGTTGGGCGCGGTAGCTCACGCC
TGTAGTCCCAGCACTTTGGAAGGCTGAGGTGGAAAGATCTCTTGAGCCCA
GGAGTTTGAGGCCACTCTGGGCAACATAGCAAGACACCGTCTCTACAAAA
AATGTGCAGGTTGAGGCTGCAGTGAGCAGAAAAACACCGCTGCACTCTAG
CCTGGATGACAGAGCGAGACCCTGTCTCGGAAAAAAAAAGAAAAAAGACA
GGGTCTCGCTGTGTCACACAGGCTGGAATGCAATGGTGCAATCATGGTTC
ACTACAGCCTGGAACTCCTGAGCTCAAGCAATTCTCCTACCTTGGCCTAC
CAAAGTTCTAGGACTACAGGTGTGAGCCACCACACGTGGCCTCAGGAGAG
ATCTTAATAATAAAAGGACAAATTGCCTTGCATCCCTTAGGGGCAGGATT
GACACATCCAAGGATCAGGCAGAAAGCCTGTGCGGAGTGGGATGAGCAAA
GAGAAAGGCTGAGAGTTGTGAAGAGGGAGATGCAGTGCCAGCTAGGACAG
GCCTTTTTGGGCTATGGGAGGTTTTCAGAGGAGACCCCACCTAAACTAAC
CCATAACATTGCAGTGGGGACCTGTTGAAGTCATGGACTACTACCTGAAA
GCCAGAGAAATGGGAGGAGCCTTTCCTCTGAGGAGGGACTCTAGTCCATA
GGTATCTTGCCACCAAATACATGGACAGGCCCTGGGGGAAGATGGTGGTA
GCCCAGCTGGAGGAAAACCATTTGCCACCTGAACTAGCCCAGGGTAAGCC
ACCCAGGCACTGAGGGTGCACACCCATGCATGCACACACAGAATCACACT
CCTTCCTATTATTCCTCAATTCAGGGGTCTCAACACCCATTTTTTTTGTT
TTTGGGGTTTTTTTTACATGTTTACATTTTATTTATTTATTATTTTGTGA
CAGGGTCCCACTCTGTTGCCCAGGCTGGAGCACAGTGCAGTCGTGCAATC
ATATTAGATTGGTGCAAAAGTAATCACGGTTTTTGTCATTAAAAGTTTTG
CCATTACTTTTAATGATAAAAACCACGATTACTTTTGACGCAACTTAAAA
GCTCACTGCAGCCTCAAAATTCCTGGTCTCAGGGAATCCTCCTGCCTCAG
CTTCCTGAATAGCTGGGACTACAGGCACATGCAATCCTACCTGGCTAATT
TTTTAAAAATTTTTTTTGTAAAGATAGAAAGTCATTTTGTTGTCCAGGCT
GGTTTCAAACTCTTGTCTTTGTGCCTCCCTCTGCCCTGTGCAAGACCTTC
TGGATGCCCACTAATGAAGACTTCCAGGGAGAGGAAAAGTAAACATAGGT
CCCTGATCAAGGGACCAGGGTTTATCGACCACAAACAGCATGCCCAGATT
CCACTGGCAGTCCTAGAGGTCGCATTTGCCCCAAGTGTGTGTGGAAGGCC
TCTCCCTAGCAGTTGGTTTATACACCAGCCACAGCACAGCATATTCTCTT
AAATTGTGAACATTTGCAAAAACTCCTTGAGGACAACTATCATGTCTTGT
GTACTTTTGTTTTGTTTCCCTTCCCCTATGTACACGCGCGCGCATGCACT
CATGCACGCACGCGCGCGCGCACACACACACACACACACCCCTCAAACTGAA
TGCCTGGTGTGCTGAATGGATGAATGGCTAATGTAAGTCATTCTAAAAGC
TACTTTCTTTGGCATACCATCACCTTTGATTTCATCTTTCTGGAACTCCT
ATGTTCCCAGATGAATTTGGAAAGCCCTCAGGAAACATTTCAAAATTGCT
ATATGGGAGAAATGGGAGGGTCTCTCTAGAAATTTACCTGCCACAGGTAT
TTCTGGTAAGACACAGCAAAGGTGGCACCACCCATTCCTCGTTACAATGT
CAATGCCAGTCACCTTCCTGTCCCATAAAACTTTATTAAAGGTGCAGAAT
TCCCATGGAAGCAGGTGGACACCATCTGCTTCCAGCCAGCCAGGGGAGCA

FIG. 4 (45 of 61)

AGGTGTCCACTGTGCCTTTGTGGCAGGAACTGCGCTTCTCTACTCTCCCA
CTTTGAGGCCTCTGGGGCTGGCCTGCTGCCTCCTCATTGACAAGGCTGCT
TACTGAGCAGTTCATTCTGAGCTGGACATAGTGCTTCTGGTGAGTCTCTA
CTTCTATTTAACCCAAAGATATTCTTTCCTAAGGAAACGCTTTCCTGTCG
GGGGAGGTTAGCTCCAGATGGAAGTCACAAGTGATGGCATGGTAGCTCTC
ATCCGTTTGGGTGGATGATATTCACGGAGCACCACCATGAGCCAGTCATG
GAGGTGAACAGTATATGCCAGCCCTGAATCAGGTGCATTGACAGCAAGGG
AGACAAGCAAACAAAGCTGAGGTTTGCTGAGGATGTTCAAGACTCACACA
GCACAGAGGAGCATCCACCACCCAGCTTGGGAAAGGACTTGTTATAGAGG
GGGTGAAGCATGAGCTGAGTCTTGAAAGACTAGAAATTAGCCAAACTACA
AGGAGGAGAAGGAGTTTCCAGTCAGGAAGAACAGGTTATGCAAAAGCACA
GAGACTAGAAAGAATATCACATTCAAGGAACTGCAAATAGACAGGAAAGA
TTGATGCGTGGGATAGGAGAGGAGGGCAGGGGATTCCAGGTGGGCCCTGC
TTGCCACACTCAGGAGCTTGAACTTATCCACAAAGGAGGTGTGGAACCAG
TAATGAATGGGTTTTGTGCAAGGGCTTCATGTCACCAGATTTGCTTTTTG
GAGATACTTCTGTGGCTGATATGTGAGGAAGGGATGGAGGAAGTTTCCGT
GGCAATCAGGAAAACCAATTAGCAGATGATTCAAATGGCCTAGGGGAAAA
GGGAGGAGGACTTGGACTACCATGCAGCAGCAGAAATGGAGAGAAATAAC
AGATCCCAGGCACTCAGGAAGCGCTCAGAATGAGCCCTTCAAAGAACTTA
TGGTAGGTGATGGATGGATGGAGTGTGAGTCCTGGGATAGCATTGCCTGG
GAAAATACTTTCTAGTTGAGACAGGGAAGTGGGCCAGCAGAAATGGAGGG
CTTCTTCTTTTTGCTTTAAATACTTTTATAATATTTGGAACTTTGAAAAT
GAGCAGATATATTAGCAAAAAGCCTAAAAGGGATATTTTTGAAATCACTG
CTAGTTCTAACATATAACTTTCAGCTTGCACACATCATCAATTAACTTTG
ATAGCGCCTTTCTGAAACTATCATCCCAAATAGCAATCCTTGTAAAAACC
TATTTTGAAAAACGGGCCTTGTAGGATAGCCTCACAGATGTTTTGTGGTA
GATTTTCTAACATTCTAATGTCAGGGAGTGAAAGGAATCCCGTTAGAAGT
TGGAAAATTCTGGAATCTCTATTCATGGTATTAAAGTTTTGCCGTCACAC
AAAAGTTTAACACCTTTACACAATCAGACTTCCTCATTTTACATTGCTCG
GTAATTAGAGGAAATCAGTCACCCAGAGCCTGGGTCCTAGACTTGACAAA
ATGCACCCAACAAATCCTGAGTGGCCTTGCTGAGGACTTCTCCCAGAAGA
TAGAAAACTCAGTTCCAGCCAACAAGGGGGAAGCAGCTGAAGAAGTGAAA
TTAACAAAGTCCTGGAAGGAAATGACCAAATCATCTTTGATTGTGTAATA
ACCAGAGAGTAGAATACAGCTACGACAGACATTTTGGGAGAGAAGCATTT
TATCATAGCTTTTAGAAGAGAATATTTTTCAGCATCATAAGCACACAATT
CCAAGACAGATACTTTCAAGGGATTGTTTTGACG
>Contig53
ATGTTNNGGTTTTGGGACCCCATTCAAACTTCATGTTGAATTTTAATCTT
CAATGTTGAGCGAGGTCCTGTGGGAGGGTGATTGGATCATGGGGGTGGGT
TCTCCCTTGCTGTTCTCAATGATAGTGAGTGAGTTCTCACAAGACCTGGT
TATTTGAAAGTGTGTAGCACCTCTCCCCTTCATTCTCTCACTCGTCACTG
CTCCGCCATAGTAAGATGTGTGTGTTTCCCCTTTGCCTTCCGCCATGATT
GTAAGTTTCCTGAAGCCTCCCAGCTATGCTTCCTGTACAGCCTGTAGAAC
TGTGAATCAGTTAGACCTCTTTTCTTCATAAATTACCCAGTCTCAGGTCA
TTCTTTATAGCAGTGTGAGAGTGGATGAATATAGTGCCATATGTTTGTAT
TCCCAGCTACCCAGGAGGCTGAGGTAAGAGGATTGCTTGAGCCTGGGAGT
TTAAGGCTGCAGTGAGCCATGACTGTACCACTGCTCTCCAGCCTGGGTGA
CAGCGAGACCTTGTTTCCAAAAAAAAAAAACCCAAACTGTGTAAAATGTG
TTCATAAAAGTGTCTTGCTCCCACACCTGTCCCTATATATCTTATTCCTC
AGCCTCCGACAACTACTTTATTCATTTCTTATGTATCTTCCAGAATCAAA
AAAAAAAAATCAAATACAAGCACAGTGGAATGTATTGCCCTTCTTCCCCT
CCCTTTTGTTACATCAGAGTTAGCATATCATAAATACGGTCTGCATTTTC
TTCTTTTTCAGCTATCAGCATGTTTTGGAGAGGATTTCATATTCGTGCAG
ACAGCATGTATTAGTCAGTCCTTGCATTGCTATAAGGAAATACCTGAGAC
TGCATAATTTATAAAGAAAAGAGGTTTAATTGGCTCACAGCTTCGCAGGC
TGTTCCACAGGAAGCATGGCAGCATCTGCTTCTGGGGAGGCCTTAGGAAG
CTTTTACTCATGCAGAAGACAAAGCGGGAGTGGATGTCTTATATGGCAGG
AGCAGGACTGAGAGAGAGAGAGAGAGAGAGAGAAAGGATGCCACATACTTTT
AAACAACCAGATCTTGTGGGAACTCTGTCACGAGAACAGCACCAAAGGGA
TAGTGCTAAACCATTCATAAGAACTCCACCCCCATGATCCAATCACCCCA

FIG. 4 (46 of 61)

CACCAGGCCCCACCTCCAACATCGGGGATTACAATTTGACATGAGATTTG
GGCTGGGACACAGAACCAAACAATACCAGAGTGCTTTCTCATTCTTTTCT
ATAGCTGCCTAGTATTCTATGTCCTTTACTTCATTTAGGCAGTCTCTTGT
TGATAGACACTTGGGTTACTTCCAATTTTTCCTATTACAAATGATGTGCA
ATGAATAATTTTGATCATTTTCCATTTCACATGGGTTATGTCCATCTGTG
GGATAAATCTCCAGGAGTGAAATTGCTGGATCAAAGGGGAAGTGCACTTG
TGATTTTCATAGTTAGCAAATTTTGTTCTATAAGGGTCATATCAATTTAT
AGTCCCACGCGTAATATTTAACAGTGGGGATTTCCCGACAGTTTGACCAA
CAAGGTCTGTTGTTAAACTTTTGATTTTTGTCAATCTGATGGGAAAATAC
TAGTATCTCAAAGTGCTTTTAATTTGACTTTCTTATTACAATGTTAAGCA
TCATTTTACTCTGCCCAAGATCAAATAGTATTTTCTTTTCTGTGAACAGA
CTGTTAAGATCCCTTGCCTCTTGTTTTGCTGGATTTTTGTTCTTTTTTTT
CAAATGTTTTGAGGCAGTTCTTTACATGTGAAACAAGTTATCTCTTTATC
TGGGGTGTGAGTTACAACTACTTTTCCTCTGGCTTGTTTTGCGCTTTGAC
TTTGCTTCTGGTGATTCCCGCAATTCTGAAAGTGTACTTTTTGCATCATT
CATTCTTATACACCCATGCTCTTGTTCACGCTGGTTCCTCTACCTGAGGG
CTTTTTCTTTTCTTTTCTATCTGGGAACATTTTTTAGAGACAGGGTCTCA
CTCTGTCATCCACGCTGGAGTGCAATGGTGCGATCACAGCTCACTGCAGT
CTTGAACTTCTGGGCTCAAGCAATCCTCCAGTGTCAGCTTCCCAAGTAGC
TAGGACTACAGGTGCATGCCAGCATGCCTGGCTGATTGTTTTATTTATTT
ATTTATTTTTTGTAGAGATGGGAGTCTCACTATGTTGCCCAGGCTGGTCT
TGAACTCCTGGGCTCAAGCGATCTTTCTGCCCCTGCCACCCAAAGTGCTG
GGATTACAGGCGTAAGCCACCATGCCCAGCCCATGTGTGGAAATCTTCTG
TTTATCCCTTTAGGCTTGATTCTTATGTCGTTCTCCTCCCTCCTTCCTGG
CTACTCCTCTTGTTCTTTATCTTACTCTACTTGTCATGTTACCTTGTTTC
TGCTTATAACTAGCTGCCTCTCCTATCTGAGGAGGGACTTGTGACTGTTC
TCATCTCTGTACTCCCAGGTCCTAGTACATAGCGCTTGCTCAACAGATGT
TTGGTGCATTGATAGATAAATCAATGGTAGCTGTTAATACCAGTCCTGAC
TCCCTGCAGTGCTTCAGCTGATCCTGTTCCAGATGTGCACTGAATATCTT
TCTGTTGAACAACAGAAATAAAGGGGATGGGTGAGGAGGATAGTCTTCGG
TGGCCAAGGATATTTGTAGGTACTTTGCAGCACTCAGCAATGAGGAGTGG
GCTTTAGTCCCCCAAGAACTCTCACAGCCCTGTTTGTCTTTACTGTTCAG
TGTCAAATCCAAGACAAGTCAATGATCAGGAAAGACCTTTTTTTTTTCTTC
AGTGAAGTTTATTTCAGAACCATTGAACAGTATGATATTTGCTCATTTAT
AAATATTCCCATTTAAATAATCTGAGCTTATATATTTTCAGTCTTAATTA
AAGGACTTGATTTAAAGAGAGCACACCAGTCCAAATTGAATTGATTCCAT
AGCTATTAAAAACTAGGCTCTTTTACAGACACTGCTACTTCTTGCCCCCT
TTGAATAAATTAGACCAATGAATAAAACAAACAAACAAATAAATAAATAA
ATAGGGAAGCGGTTGCTCATCAGAATGTGGGAGCGAATGACAGAGGGTTT
CTTAGAACCAAATGTGGCCGTGGTTTCTGTCAGGCGGGCTTTAAGTGAGT
AGGAGAGGTGAGAGAGGCCTGGCTCAACAAAAGGGCTGGGGATTGGCCCT
GAAAGGAGAGAGCTGACTGTCCTGGCTGATGGACAGGAGATCCTCTTAGC
ACTACCCTAAGGCAGGCAGTTGGGCATTGGTGTAGACAACAGGAAAGTCC
AGGCTATAGCCGTACTCAAAAACCTTTCTGTTCCCTTTCTGCCAGCCCTA
GGGATTGAGTCCACATTCAGCACAGGACTCTCTGGGTACAGCTCTCTTTA
GGAAGACACAAATTGCATGGTGAAGTCAGTTATATCCTGGCCGCCTTTGG
TCCCTCCCAGGAAGACGGGCATGTTTTCTGCTTGAGAGGTGCTGATGTAC
CAGTTGGGGAACTGGGCAGACTCAAATTCCAGCTTGTTATTGATTTCTAT
CTTGTTGAAGACAAATCGCTTTTCCATCTTCTTCTTTGGGTAATTTTTGG
GATCTACACTCTGCAGCGAAAGAGAAAGAAGAATTTTTGTGGGGCAAGGG
ACAAAAATGCTATGGGAAAGATGTTCTTTGGGTTGGCCAGAAAGGAAACT
GACGAGCAGGTCACATGATCAGGAGCCACACTCCTGAGTTGTAACTGGGC
CCCCAACTTTCTGTGTGATTATTAAAAGAGCCCTTCTTCTTTTCTAAAAC
TTAGTGCCAAATGCTGAGGAGCATAATGTAGGTGAGAATTTTTTTTTTTT
GGGGGGGTGAAAATTAAGCTAGAGCTTCTTGAAGTACCTAGTTTCCAGGG
GCTTTTTATTGTATTTTTCCTTATGGTCCTTGAAGTACCTAGTTTCCAGGG
ATGAAGCTTTTGCTGAGAAAGCTGGAGGTGATAGTGGTGGTGATTTTGGG
AGTGGAGTGGACGTGATAATGGGACCCCTTTAAGTCATCTATTTCCCAAGG
TGTCTATCAAATGAGAGCAGCCCTAACAATATATAATCTGTTGGGGTTGT
AACTATGGTAGGACATAATAACATCGGCAAAATGATTTAATTTTCTGCAG

FIG. 4 (47 of 61)

101/118

CAGGATTGAAGGTTGCAGCAGTTAAAATTATGTTAAATTTATTTACAT
TAATGCAAAATTGTCAAATAGACCTGTTCCCAGCTTTTCCTAGGGATGGG
GGCGGGGAGAAGGTGGTTGTCTGGGAATAAGTGGTAGCAGGAGGCTGAGA
AGGGCTTCATTCCATAGCATTCACTTACCTCCAGCTGTAGAGTGGGCTTA
TCATCTTTCAACACGCAGGACAGGTACAGATTCTTTTCCTTGAGGCCCAA
GGCCACAGGTATTTTGTCATTACTTTCTTCTCCTTGTACAAAGGACATGG
AGAACACCACTGAAGAAAGAAGGGGGTCTTGTGGTTAGGGACACAGCAGT
GCAGGGTCACCCCAACCCCTAGGCCCCATGAGTAGGATACATGTAATTTG
GTAGCCTCTGTGGGAACCCACAGTGAGGTTCCTTGGCCTAAGACACAGGA
TAACTTGACTTCTCACAGACAATAGCAGGGTCATTTTGTTGATTTAGGGT
TTCCCCTCAAAGGCCTGAGGGTTTCTCAGAGCCTCATAGCAGTAGGAACG
GAGAATGAAAGAGGGTCTACATTTTAAATGCTGAAGGAAGGAAGGAAGGA
AGCCATTGTGTCACTGGCTGGCAATGTGCCCATCCACAGGAGCGGAACAA
CTTGATCAATGTGGAAGGAAAGGAAAGAGGTGAGGCTGTACTTCTGCCAG
AAATCAGGCACCAGAACTGTTTCAGGAACAGAGAGTAGCCCATGGGAAGA
AACTGGGAGAGGAGAGGCTGAGCTGGGAAAGTGGCTCCAAAGAGAGACAC
TCATTTTGATCTTCCTCAGTCACAGCAGTGTCAATTGGAAGGCCCTGGGA
TCACTCTTACTACCCGATTCCAAAGAAACAGGATTTTCTTGGCCTGGCTG
AGAGCAAATAGCTTCCCCCTTGAGTGAGGCTGTCCTTCAAAGTCAGCAGC
CTTAGTTGCCCACACTCCTGTGCAGAGGCTTTGGCTACTGTGGCACGATG
CCAGGCAGATCACCACAGCTAATGATGGGTTCACCGCACTTGAAACTTTT
GCCCGTTACAGCGGAGAGATATAAGTTCCTGCTGGGCGGTAAAATTTCCC
TACAAGGAACCACCTGGCATTGGGTGGGACGGATGTTGGGGCAAGGGGGG
AAGACTGGGGAGGGGGATGGACACATTATCGCTCCAGCACTCTTGTTTCA
GCCTCAACAACAGGAAGAGAGAACCCACAGGCAGTTAGGCCATGTCCATC
AAATGACCCCATATTGTGGAAGAATTGACATTGCACTATGCCCAAGAGAC
TTGGGTGGACATGGTCCTGGGAGTGCTTGAGCCGTCTAATTTCTCAGGGT
CACACTCCTGTTAACAAATGCACTGGCCAGTGCAATCAAATGTGCCATTT
CTAGGACCAAAGTTTGTATATTCCTTTTTAATATTTTTTTTCACTTGTGT
TGATCATTTGCCTTAAATTAACTTTCTACTTTGTTTAAAACATGGAGAAT
TAGCAAGCTGCCAGGAAGCCAGGCAGGGAAACCAGGATGTTTCCATTTAC
CTTGTTGCTCCATATCCTGTCCCTGGAGGTGGAGAGCTTTCAGTTCATAT
GGACCAGACATCACCAAGCTTTTTTGCTGTGAGTCCCGGAGCGTGCAGTT
CAGTGATCGTACAGGTGCATCGTGCACATAAGCCTCGTTATCCCATGTGT
CGAAGAAGATAGGTTCTGAAATGTGGAGCACATGTTGTTTAGGTATAAAA
TCAGAAGGGCAGGCCTCGTGAGGCAAGGTGGCAAAATTTGATTTCTTGGA
GGACACCTGAGCATATACGGTCAAAGTCTGATGACAACACCAGTAGGGAT
GAAGCTGGGAGTGGGGTGGCTAAGAACACTGGACCTGACACTATTAGACA
TGGGTTCCAGCTTCAGGTCTATTACTGCTCACTGTGGCCGAGCAACAGAG
CTACTTAGGTAAAATGGTGATGGTCATAACACTAGCCCACAGGGAGGTTA
CGAACCTCTGGTGACAATGTAAGTGAAAGGCCCCTGAGAAAGAGTGAGGG
AGTTGCAAATGTCAGTAGCCATCAAGATCTTCTTTAAGAATAGTTTCCAC
TAAAGAGATGATTGCTTTGGTTTCCAGCCTTCTTTGTTTTGTCTCCCCGC
TGGGCCTTCTACCTTTAAAGGGCTTTGGCTCTGGGGGAATTGAGTTGGCT
GGGGCTTGATGACTTCCAAGAGGACACAAGTGGAGATCTACTGCCTGCTC
TTGGCTAACTACCTTCTTCAAAGATGAAGGGAAAGAAGGTGCTCAGGTCA
TTCTCCTGGAAGGTCTGTGGGCAGGGAACCAGCATCTTCCTCAGCTTGTC
CATGGCCACAACAACTGACGCGGCCTGCCTGAAGCCCTTGCTGTAGTGGT
GGTCGGAGATTCGTAGCTGGATGCCGCCATCCAGAGGGCAGAGGTCCAGG
TCCTGGAAGGAGCACTGCGGAGAGAGCGAGGGAGGGAGCCTGGTGAGGTG
GTCCTGCCAGGAACCATGCTTTGACATCAGAGAGTAGAAAGCTCAGAGAG
GAGGAAAGGGCTTGAAAGAATCCCGAGCTTCTAAAGATCATCCCTCTCTG
GGCCAGGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAAGCCGA
GGTGGATGAATCATTTAGGTCAGGACTTCAAAACCAGCCTGGCCAACATG
GCGAAACCCCTTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGG
GTGCACCTGTAATCCTAGCTATTCAGGAGACTGAGGAAGGAGAATCGCTT
GAACTCAGGAGGTGGAGGATGCAGTAAGCCAAGATTGTACCACTGCACTC
CAGCCTGGGCAACAGAGTGAGACTCTGTCTCATAAAACAAAACAAAACAA
AACAAAACAAAATAAAATAAAATAAAATAAAAAGATTATCCCTCTCTGAA
GCTCAAGGAGGTTAAGGGTGTACTCAAGGGCACACAGCAGGTTAGAGGCA

FIG. 4 (48 of 61)

T62/118

GACTCAAGACTAGAATGTGGGCTTTCTGACACCTTACAGGCTATTCTTTT
AGAATAAATCCCATTTCTACTTTGTTCATCTTTTTGTACATGCCCCACC
TACACCATACATGTATACCTTCTCTATATCTTTTTGTATCCCTAATGCTG
TCACACTATGATTTGCTTTTTCATGCAGATGACCATAACATTTTCCATTC
ACCTATGCTCACTCAGCAAGTATTCAATTTTTCTACACTGTTCTTTTTTT
TCCTTTTTCATAACACTGTCTCATAGGCATTCTGCAAATCCTGTGAGAGT
ACTTTTTGTGAAATGTTACCACTTTCCTCTTATTCAGAGAAGCTCCGTAT
TAAGGCTTCACTGAGGTTGCCTTAAGGCATGATAATGGTTCAAAGGCTTG
AAAGACAGTTAAAGAGACCTGTAAGTGCACAAAAGAAAGTTGAGCAGGAG
AGAATTTCTTGCCTGGAGCAGAGCCAAGCTACTGGAAGAGGCAATGGGGG
CAAAGGCCAGGCAGACAAGCCAATGGGCTCCTCCCACAGCTGCAGCCAAC
AAGTTATGCCAGTCTTAAAACTTCTAAAGAAATATGTTTTTAACAAGATT
GAGGACTGGATTATGAGGCTAGGGGAGGCTATCACAAACTGGAATAAAAT
AAAGCCAGAGAAAAGTGGCTGCCTTCCAACCTGCACAACTGACCTAGCTA
GGCTGATGGCTGGGCCACCTAGGAAGGCTACTGAGCATCATATAAAACAG
AAGGGACAGCAGGAATATAACATGGCTCTTTGTAAGGATGAGTCTGAAAA
ATGACCATTTGCTGCCCAAATGCCCTTAGCTACAACTGAAAATATTTCAG
AACTGGAGGTTGCAGGATGCTGGAATCTCAGAGATCATCCAGCTCAGCCC
TTTATTTTTCAGATGAGGTCCAAAGCGGGTAAAATGACTTGTCAAGGTCA
AACAGCAAGTGAATGGTTTTCTTTCAAGTCTCAATTCATCTTTTTGTTTA
TATCATCTATGTCTTGTTGTTATAAGCTTCACCCCAGGTAGCAAAAAACT
ATTCTACTCAAAAGGGGTAGACATATGTTAGTTCTCAAGATCATCTCTTG
GTTTCAGAGTTTAACTCAAGTGATTGGCATAGGCTGAATCCATCTCTTAA
AAGGATAATCAAATTTATGTTGAAGACTTGGTTGTCTTCCTACTATGAAA
TGGGAAACATTATCACTACTCCTCCCCTGTCACCACCAAGTGTGGCCACC
ACCACCAACGTTAGTGAGTGACTGTGGTGATATGATGACCAAGTGGCCAG
GTCAGCAAGTGGTGCAGCCTGTGTCTCACTGGAAGAGGTTAAAGTCTTTC
TAAAACAAAATACCATGGCATCAAAGTGGCCCAGAACTCCCTTCTTTGAG
CTTTCCCTGTGTTAGAGCCCTTCCTTGGGTTGGGAGTTAAACCCATAGTC
TTACCTTCATCTGTTTAGGGCCATCAGCTTCAAAGAACAAGTCATCCTCA
TTGCCACTGTAATAAAAACAGGGACATGTCTCAATTATGTCTTCTAAACA
GGTTTATTTTTCCTTCCCTGTGTACAAGACTTGACTGTTCATAAGAAACT
GCAAACAGCCTGCCTCTCAAAGCTGCCTGAAACACCTGGCAAGTTTCACA
GTGATATGCGCAGAACAGTCCAGAAGGCAGATTCTAGGCCTGGCAGGTGG
GCACCCTGGGTGCTCCCTGTTGGATCTTGAGGCCTAACCTCTAGCCCAGC
AGAGTCAGCTAAAATCTGAGCTCTCCCTCTCCCTCCAAGCCACACTTTGC
AAAGGGATTCCTTGTATTGTGGGCTTGGAATCTTTTCTCCCCATTTGCCT
CTGCAGGAAGCCCTTGCAACAACACATCTGGATAGCCTCCAGGTCCCAAG
GCTGGAGGGACTTGTAATGGGAAAGTAGTCTTTAAATCAGATTTACTTGG
CACCCTGTTTGCCACTGAAAGAGGCAATTTAGGGGAAAAATCTGGTCTCC
AAGCACAGATAACACTCTACTCTTGAAAGAGGAGACCTGCTCATGTTACT
GGTCTCAGCGTCTCCACTGACCTGTAATAAGCCATCATTTCACTGGCGAG
CTCAGGTACTTCTGCCATGGCTGCTTCAGACACCTGTGTAAAAAGGAGAA
AATGAGTGACTTCCCCATGACGGCTACGTTCATGTGTGATTTCTCTCAGC
ATCCAGTGCATGGCAGTCATGCAAAGAAATGATCTCTGAGTAAATGAATG
AATGTGTGAAAGAGAAGTCCTTTGGGTCTAGAGAAAAGCATTTGCTAAAC
CAAACCCCAACTAGCAATGTATTGGCTAGGAGAGCTGGAGCAGAGGCTTT
GACACTAACCTTTAGGGTGTCAGCTGTTAGATAAGCAGTATCCATTCCCA
GAATATTTCCCGAGTCATAAGCATTATATTACACCTGGCATTTTTGCAAA
AAGCTGAGAGAGGGAGGCAGAGAGGGAAGGAGAGGGAGAGACAGAGAAAG
AAAGAGAGAGAGAGAGAGAATATGCATACACACAAAGAGGCAGAGAGACA
GAGAGACTCCCTTAGCACCTAGTTGTAAGGAAGATTAAAGTCATACTTGA
GCAATGAAGATTGGCTGAAGAGAATCCCAGAGCAGCCTGTTGTGCCTTGT
GCCTCGAAGAGGTTTGGTATCTGCCAGTTTCTCCCTCGCTGTTTTTATAG
CTTTCAAAAGCAGAAGTAGGAGGCTGAGAAATTTCTCTGTTGAATACCTG
ATTTCACAATCAAGTTAAAGGAAAGGGGAAAAGAGTATTGGTGGAAGCTT
CTTAGGGGAGGGGACTAATAAACTGAGATAATTCTCTGGTTCATGGAAGG
GCAAGGAGTAGCAAACTATGACACATTTTGCAAATGTATCACCATGCAAA
TATGCATTGTTTTCCTGACAATCGTTGTGCAGTTGATGTCCACATTAAAA
TACTGGATTTTCCCACGTTAGAAGAATGTTTAAATTTAGTATATGTGGGA

FIG. 4 (49 of 61)

CAAAGTGGAAGACACACAGATTTATACA.GCACATACTTTTCTTCATTCA
CTTCTTTGTACTTAAGTTTAGGAATCTTCCCACTTACAGATGGATAAATG
GGTACAATGAAGGGCCAATAGCCCTCCCTGTCTGTATTGAGGGTGTGGGT
CTCTACCTTGGGTGCTGTTCTCTGCCTCGGGAGCTCTCTGTCAATTGCAG
GAGCCTCTGAGGAGAAAATTGACCTTTCTTGGCTGGGGCAGAGAACATAC
GGTATGCAGGGTTCAGGCTCCTGACGGAGTTGGGGCAACCCTGGAGATAA
GCTCACACAACCCTGCAAGACCAGGTGCTGTTACCCTAGCCAATCTCATG
GATGAACCAGATCAATGCCAGATGAGCTCTGCCTAAAATGATTTTTTGGT
GAACTCTGAAAAGTGGAATATTGTTTCTGTAAGAATATCCATCTGAGACT
CTATCTCTTGGTAATACCAAGAGTTATCAGTTTCTCTTTAACCGAGACAC
CAGCAAAGTGCCTGCTCCAGGGTACTGCCCAGGGGAGCCCTCCATTTGTA
GAATGAATGAGAGTCCAGGTTATGAACAGTGCCTGGAGTGTAGGAACACC
CTCCTTTGCCTCTTTGACAGGTCTGCATCATAACACTTTTTTTTTTTTTT
TGAGACAGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAGTGGCACGATC
TCGGCCCCCTGCAAGTTCCGCCTCCCGGGTTCACACCATTCTCCTGCCTC
AGCCTCCCCAGCAGCTGGGACTACAGGCACCTGCCGCCACGGCCGGCTAA
TTTTTTGTATTTTTAGTAGAGACAGGGTTTCACCATGTTAGCCAGGATGG
TCTCGATCTCCTGACCTTGTGATCTGCCCGCCTCGGCCTCCCAAAGTGTT
GGGATTACAGGCGTGAGCCACCGTGTCCAGCCTGTAACACTTCTTATAGC
ACTGAGTTGAAACCTTGCTCCTCCTGGTTCCTCCAGGAAACTGAAATCTT
TTTGAGCCAAGTCTAGCACAGTGCCTGGCATGTACATTCAGGTGGTAGAG
TTTGCTGCTTGAATGGGTGAATGGGAATTTGACAGCATTTTTATTCAAAT
TAGTATGTGCCAGGTATCGTGCTCGCTCTGCATTATCCAAGGGAGTGAGC
CTCTGTGCAAGTATTTGAGACACGAGGGAAATAGGTTCTACTGTGGGAAA
AAGAGCATTTCATGGACTTGCTCTCCAAGCAGCCTTCTGATTTTTAATTT
GGCTCCCAGTATCTTGATATCAGGAGTCAGTCACAAGAACTCCATCTTTA
GTAAGTTATATTTTCCACAGGAAATCTAAAAGCTGTTCAACATGTTAGTT
TCCTGTGAATTTGATAAGCCATAATCCATTCCTAACACTGAGCCCTCCTG
AAATTTGGTGTCTGGTCCTGCAGATAGCTAAAAGCCCTGTCTGGGTGGCC
TAGGGGACTCCTCTGTTTTGCCTCCACAGGATCCACTTTGCAAATTAACC
ACTGGTTCTCCCGTTGTAGGAACTGCCACCTTCCTCAGAGCCTGTCTTTC
TTCCTTCCTTCCTTCCTTCCTCTTTCTTTTTCTTTCTCTCTCTCTTTCTT
TCTTTTCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTT
TCTTTCTTTCTTCCTTTCTTTCTCTTTCTCTCTTTCTCTCTTTCTCTTTC
TTTCTTTCTCTCTCCCTCCCTCCCTCTCTCTCTTTCTTTCTTTTTCTTTC
TTTTCTCTTTTCTTTCTCTCTTTCTTTCTCCCTCCCTCTCTCTCTTTTTC
TTTGTCTCTCCCTCCCTTCTCTCTCTCTTTCTCTTTCTCTCTCTCTCTCT
CCTAGACAGGATCTACCTTTATCCCCCAGGCTGGAGTGCAGTGGTACAAT
CATGCATTCATTGCATGATCACAGCAGCCTCAAACCCTTCCTCAGAGTCT
TTATGCGGCAACCAGCAGGGTCTGGAGGGTTGGTGGCTCTGTGAACTCTC
CTGACAGAACACAGAGATGTCTTTGGTCTGTTGATGTGATTACAAGCTGA
ACGAAGGAGGATCAAAGCCAGTGACAGGAAGGGAGATATGCAAGGGACCC
GAGCATCAGCTCTGAGTTAGTCCATTCTGCTTCTGGGACTTGGGATACAG
GTCAGAAACCTTGAGCTTCTACTTCTCCATCTTCCAATTGTAGCATCCAG
GACCTCAGAATCTGCCAGCTAAGAGGAGCCCGTAATGATTGTCTGGTGGGA
TATGGTGGGACCACAGAGATGAAGACATGAATAGCTATTTGAATGTGAAC
AGCAGACGAAGAAATCAAGGCTAGGAGGGTGGAAGTGACTCATCCAATAG
CACAGTGTGGTTGAAGCAGCACTAGTATCCAGGTTGCATGAGCCCCTGAT
GCTTTCGCTCGAGGGAAATTTTGGAGCCATGGGGCAATGCCCCCTGACGT
AACAGTCTCCACAGTTCTGCCATGTCTCATCCTGGCCCTGTAACCTGGAC
CCAAATCTGCTACCATCCCATCCATCTCAGGAAGTGAAACCTCTTATGTC
AAATAGGTTGTGCAACGTATGTATCAGATCCTGTCTTCCCAAGGAGACCG
CTCAGGCCACAGCACTTCCTTCCGATCCCCAATGAGCAGAAAATATCTCG
CTATAAACATAGTTGGCACTAAGGGAGGGAGTGGAAGAGTGATGATGATG
TAGATGGTGATGTAGCCCCAAGGAAGTGGAACAAGCAGAGATGGGGAGCT
GGAAATGCCAGGATGCTCCAGCTTTTGGGGAATTATTCAGCTCTTGAGTC
ACTAAAGCCTTTCTCAGCTGCAAGTTCCTCTTTACCCTGTCAGGTCATTC
TTCCAAGACAGGAGACTGACATTTATTCAAAGCAGCAAGTGCCCTGATAC
CATCTTGTGTCTAATCATGGGCTTCGCAGCCAGTTATCAAGGTTGATCTC
ATCTCATTGGTCTTCAATCATTTTGAACAAGAAGACAAGCAAAATAATCA

FIG. 4 (50 of 61)

TGGGTTAGTTCTTATATTATTGTGTGTACATGCAGTGATGTCTGTTCTTT
GTAGTGAGCTGTTCCTTCCTTGTTCACCCTCTTGCTTAGAACAGAACTAA
GCAATCTGCCCCCAACATTTTCCCCAATTTCCCATCTCATTCTTGGCACT
GGCTTCCTAATATTTGTTCTTATGAGTCATTTTCTTGTATCATTTCCATG
AGTCCCTCTGGATCTTAAAGTATGAAAAATGTTGTGTGTACCCACACCT
GTCTTTGTGGATATTTCTCTCCTTTCCCTTCTGCTTCTGGGATTATTTGG
GAATGGGCACTATGATTTTTATCATATCGCTTCCACTTCCTTTATGGCAT
CATCTCCAATGGGCTTCTTCTCCCTCTTGGATCCAGGTTCTCAGATTGGG
GACATGCAGAGTCCAAGGAACATTCCATTCTCCTCCCTGGTCTAGAACAA
GGAGGGCTTAGATATATGAGCAGGTGGCTGGGGCTGGCGAGCTATGTAGT
CTCCAATGGCTTTTCCCTGATGTCGGAGTTGTTATGTCAGTTCTGGGAGA
CCAATAAGACCTTGTCCTTCCTTTGGATCCATCAGAAAAAGCCCCTGGGT
GGGTAAGATGGATGGCAGGGCTCTCCTACTCTATGTCTTTTCTCACACCT
AGTGGGTATAAGAGAGGGGACCACAAACAGAGGGGGCTCTGGTACCACTT
ATCCAGGGTCTGGAAACATTTTCTGTAAAGGGCCAGATAATAAATGTTTC
AGGTACAACTACTCAACCTTGCATCATTTCAGAAAAGCAGTCAGATAATA
CATAAATGAATGGGTGTGGCTGGACTTGTCCTGCGGTCCCCTGTCTTATA
TCATTGTATTATATCATTTTTTCTTACATACAAATTTAGAAGCAATACTT
AAAAAAAAAAAGCCGTCCTTTATTGAGCACCTACTAAGTGCCAGGTACCT
TTTTTTCCCTCATTATCTTATTAACTCTTCATAATAACCTTTAAAGTAGA
TAATATTGAACCATTTGACCTATGCAGAAACTGAGGTTGAGACAATAAAT
TATTTAAGACCGCACAAACAGTAAATGCTGGAACTACGACTCAAATATGG
GTTAACTGAACCAAAACCAGATCTTTATTTCTCACTTTTAATTGTTACAT
ATGTTTATTGCCTCATCTCCTGTCCACATGGTGCCCATCGGCAGACTCCT
TTCTCATTCTCAGTGATTGAGTGACATTCTAAACTACATTGGCCTGGCAG
ATTCACCTCTGTCCCCTAAATGTTTCCACATTGTCCTTTTAGGATTGAGA
TCCTCTCTGTTCCCTTGTCTTCCCTCCTTTCTTCTTCTGGCGGTGACGTG
CTGTGTGAATTTGTTTCTTTCTCCTCTCAGGGTAGTACTGGGACTTTCCA
AATCAGGGTTTTTAATGATCTCTCTTCNCTTTTCTGAATTTCTTCCTTAT
TCCCATTCACTTTCTCATCTATAAGTGGCANCTTTGTTGCTGGAAGATAT
CCCTTGTGCAGGGATTNCTCTTTAANAATTTGTCNNNACC
>Contig54
GTGATCGTCAACCTCCCACCCTGTAGGGCCTCAAGCATTGAGGACAATCA
CTGGCTGCCCATTAACCCAGAAATGTTGCCGAGACAGGAGGCCGTGGCCC
AAGTTCCTGGAATGGGGTATTATTATGTCAGCACAAAGGCCTTTGCACAA
ATGAAGGCTTTAAAAATGCAGTCCTAGTCAGGTGGAGGAGGGCTTATAGG
ATTCCCAGGAATCTGGATCATTCTCTTGAGAGCTTTCCCTTGTCTCTGTT
AAAACTCACATCGTACGGCCCAAATAACAACAAAAAAATGGATGTAAATTC
TTGAAATAACTTGTGGATGGGGGAACAAGGCCCACCCCCCCAGATCTGCCA
GAAGCTTCAGGTGAGGGTCCCAAATGCCAAAAAGTCTGGTATCAGAGAGG
ATGGCCAGTGACNTGGGGACACATGCCCTTTGCTGTGTCACTCAAGGAGC
AGCAGCTTCGGCCCCGCACAGTGACCAGGACCCTGGCTTCCCACGCTGGG
CAGGAGCTGGTGTCTGATGAAGGGAATGCCTGGCAGCACGTGCTGTCTGT
CTCCTCGTGTCAGCTTACCTGGCTTTGCTGCGAAGAGGCCACTTGCATTT
CTTTATTTTTTATATTTTTTTAATTTTTTAAATTTTTTATTTTATTTTTA
TTTTTATTTATTTATTTATTTTTAATTTTTTTTTAATTTTTTAAATTATG
CTTTAAGTTTTAGGGTACATGTGCACATTGTGCAGGTTAGTTACATACGC
ATACATGCGCCATGCTGGTGCGCTGCACCCACTAACTCGTCATCTAGCAT
TAGGTATATCTCCCAGGTTAATCCCTCCCCCCCTCCCCCCACCCCACAAC
AGTCCCCAGAATGTGATGTTCCCCTTCCTGTGTCCATGTGATCTCATTGA
ATTTCTTTAAAGGTGGAATCTCTCAGTGGGGTCTAATCTGTTCAGAAATA
TCAAAAGAGTATCCTTGGGAATGACTGGAATTCCAGAGTCATCTGGTAAT
CCTCATAAAACAACTCCTGGATGTCTCTCAGCACATCTCCCACCTTGAAC
GCAGGAGGCTGGTTCAAATGGAGGAGCATCGCTCTACTGCACTTTTTTTT
TTTTTTGGCCTAAAGTGCAAAAGGGGATACGTTTCATGTAAATAAATCAA
CTGCAAATCGCTAGTTATGCTGAGCCCTGTCCCGTGCTGTGGACACAAAG
GAACCAAAGGCTTTTCTCCCCGCCCAACACACACATAACACACACACAAA
ATCATAAAAACATACATACCCCCAACACATAACAACACACAACACACACA
CAAAATATATACACACAACACACACCAAACATGCCCACAAACCTGTGTCC
AAAAATAAATCCTACTGGTGGGTTTGTGGTCTCCCTAACTTCAAAAATGA

FIG. 4 (51 of 61)

AGCCGTGGACCTTCGCAGTGAGTGTTACAGCTCTTAAAGATGGCATGGAT
CCAAAGAGTGAGCAGTAGCAACGTTTACTGTGAAGAGCAAAAGGACAAAG
CTTCCACAACCCAGAAGGGGACCCCAGCAGGGTTGCTGGTTGGGGTGGCC
AGCTTTTACTTCCTTTTGGCCCCTCCCATGTTCTGTTTCCATCCTATCAG
AGTGCCCTTTTTTCAATCCTCCCTGTGATTGGCTACTTTTAGAATCCTGC
TGATTGGTGCATTTTACAGAGTGCTGATTGGTGCGTTTTACAATCCCCTT
GTAAGACAGAAAAGTTCCTGATTGGTGTGTTTTACAATCCTCTTGTAAGA
CAGAAAAGTTCCCCAAGTCCCCACTGGACCCAGGAAGTCCACCTGGCCTC
ACCTTTCAACTCCATAATGGCATGAAAATACATATGTTGTACAAAACATA
CATACACAAAGTATACATGCATCTCCCCAAATATACACATACCACAGAAA
CATACACACAGGAACTCAGCTACCTGTCAAAAGTCTGCATGGTGATTGCC
TCTGCAGTGAGTAGTTAGAAAAGTGAATTTGTTTTTCAATAAATTGGAGT
CCTTAAAAATCGTTGTAAGATAGAAAATTTTTAAAAGTATATAAAATAAA
ATATGTATGTCCTTTGGTCTAGCATTTACACATGTAGGAATTTATCCTAG
TGGAGTAATCAATGATATATGCAAAGATTTGGACAAGCATATTAAGCACA
GAATTATGTATGCATATGTGTGTGTATATATATATATATATCTGATACATAT
AATAATGTAAAAGTGAAAATAACTCAGATGTTCAAAATTGAGGATTAGTT
AGACTATGATCTGTCCATATGTGACATACAAGTTAGCTGCCCCTTATTCT
CTCGAGCTTCAACCTCCTATAAACAGTGTCCCTTGTATATCAGTATTGGT
ACAGATAATCGAACTTATTGAGGTTTTACATGGGGCAATAAAGGCAAGAG
TTTATGAATACTCCATACTACACTAGGTAGCACCCCCTATTAAAGACAAA
CTCTTCTCTCTCATTTCCCTTCCTTTCCGGAACCACTTGGTTGAATCTCT
ACAAGTCTCTATTGCAACTGCCTCAACATGGCACCCTCCCTGCATCTCCA
TCTTCCCTGTCCTGAGAGCAATGGCCTGCTGCCCCCACACTCACATCCTC
ATTCATTCCAGAAGTGAGCACCACAGAAGTGCCTACAGTTACCCCAACCA
CCTTCTTAGAAGATAAGTTAGTGTTTGTTTTGACTTTTAAAATTTTTAC
TTCCTCTTTTCCTTCACAATCTCATCCCATCCCAAGAGGTTTATCAAGAA
GTTCTCTAAAGATATGTGTCTCCTTATGGAATTTAACAGAAATCAGGGAT
TTGTATTCTAGCCATCAAGGGAATAACATTTTTCCAGGTCTTTAGACAAA
TAATGGAATACCTTGCAGTAATTAGATACACTATTGTAGAAAAGTATTGA
TGAAATGGAACGATGTTTGAGATATCATATTGAGTAGAAAAGGCAAGATA
CATTAAGTAGGAAATGTATCTTACAAAATAATTTGTCAGACACACTCCTA
TATTTGTATGTTATATAAATGCGTATGTGAAGAAAGGCTAGAGGATGAGA
CCACAGTCTTCGGTGAAGTTTAAGAGATGAGGCTGCAGCATGCTCAGAAA
GGCCTGGGTTATAGTTCTTCCAGTAATTAAGGATGTGATCTTGGGTAAAT
TGTCCATCCTCTCTAAACTGCACCACCTTTTGTCTGTAAAACAGGAAGGA
TGGTATTTACCCCCAGGGTCATCAAAGGATTTGGTTGGAGAAAAATAAAT
AAATGGGCTGAGCCCAGACCTGGCACAGTGAGAGCACAGTGGTTGACTAT
TGTGCTGGCCTGTTGTTCCTGTGTTATTGACATGCTGCTGGTGGTGGTCC
AGAAGCTATTACCTTAATTGGTTATGTGGATTTCCCCTCATACTGAGCAG
CTGTGTGTGGTGTTGTAAAACATAGCCATACACAGTAACTGACAAGGGCA
AATGTGATGGAAAAAATGCAAGGAAGTGCAGATAAATAGCTAATGGGCTGT
AGAAGGAAGCTAGTCCTTGGAGGGCTTGATCAAGGAAGGTCCTTTTGCAT
GTCACCTTTGAAGAAGAGGGGACATAGAAGAGGTATAGTGCATCCCGGAG
TGTACCTGGAAGGGAACATGAAAAGAGGACATTTTTCTCTGGGACATGGG
GACTCCACTTGCATGAACTCTGGAATTGGGGCAAAGAACCATCATGAGAA
CAAGGGCTTCCTTGAACCTCCCAGGCTCATTGGCTGATCTAAACCCTGTG
TCCCCTCTTTCCTTCACTCTCCTCTGTTTTCTATACCTGTATTATTGGAC
TGGACTGGAAGCCACCTGATCTATCACAAGTACCTTGAAATGTGTTGAAT
AGGTGTGGCACAGTCCTTAGCAGAGTGGCACTACCCCCACAGGAATTTGT
TTATACCTTTGGCATGGAAAATAGCAGGAAATGAGTGATCACTGATAACT
GAGGATGCTATTTATTATTGGCCAAAGGAATACTTGTGTTGTATTTGCAT
AACCACTCACAAACTGTTGATTACAAATGAGTACCAGACCTAGCTCCTTC
AAGTAAAGGATCCTGAGAACTGAAGGCAAACAGAGCTCCAGGAGTCCAAG
ACAGAGCCACAGACCACGAGGATCCCTGGCCCAGGTAGGTGGTCCTCCTG
CACTGGCTTTCAAGGCCAACAGGATGGATGGGGAAGTAGAGTAGCATCTG
GCCATCTAGACCCTTGCTTTTTATCCCCACTGGAAGCACATCTGAATTTC
TAAATATGATCTCTGAGACCTGCCCAGAACACCTTGCTCTCAGCCCCAGT
AGCAGCCTGCTCTCTCCCAGGAGGGCTTCCACTAACAAGTAGGGCATTGC
TGGAGGGCCAGGCAGACACTAGCTTAGGAAATCCACCAACCCTGGAAATG

FIG. 4 (52 of 61)

106/118

CTAGTCCCTTCTCTGAAGGCTCAGAAGACTGACTTTAGAGTCTAGAAAAT
ATTGGTCCTTGGGAACAGATTTTGAGTGCAAAGAGATGGACTTCAGATGG
CCAGATGCACTGCTTCTTTAGGGAATTCTGTGAAAGCTCCCTGCATTTAT
CTTAATACAGGCAGCAGATTTCATGAGTACCCCCGAGGGATGGCCCCAGG
TCCTCCAGCCTGTGAGCATCCTTCTGTCCTTCAGCAGCACCACAGTATCT
TTATATGTCTTTGGATACCTACGTTTCTGCCAGACATCTCTTGCTCTGAT
GTTCTGGCTGCCAAATTCTCTGTCAAGCGCCTCCAATTTTTTGTGTCCTT
TGATTTACCCCAACATGACAAAGGCAGTTGTGCTTCATGTATTCAGGGAT
ACTGCCAAACCACAAACAGGTTAAAATCAAATAGCAGATATCCCTGTTCC
TAAAGACCCATCAGCTCTACCCACCTGCTCCTGCTCACCGTCCTTATTGT
TGAGTCCTGAAGCCCTTCCTTGTCATTTTTATTTTTTGCATGAACAATTT
AGTTCCCTTTGTCTCACTCCTAAACCTTTCTCAAAGGATTGGATTTGTAC
ACAAACTGCCTATCTCTGCAATCTTAGAAGTGATATGATTCTGAACAAAT
CACTTAACTTTTGATTTTTTATTGGTAAGATGGGAATACCAATTTTTGCT
CCACTTCTGTCCTATGTTGGCCTGGGCTGATGTTGAAAGCTCTCGGTCAA
CTGAGATAGGGTGTGCAGAATTTATATATATAAATATATCTCCTCCAACC
CCTCCCAATGAAGCAAGTCACGTGAGTCAATCCTACCCTAAGATATTAGG
GATTGAGCCTCCTGGGACATTTGGTGGCTTAGGTTTTCATGAAAAGAGGT
TGCAGAGCAACTGCTTTTTGTTAGGCAAAGATTAGGCTACTGCAGAGACT
CAGCAAACTTCTATAGAAGGTGTCAGATGGTAAGTATTTTAGGCTTTGCT
TGCCAGATGATCTCTCAACTAGTTAACCATGCTATTGTAGCCTCGAAGCA
GCCAGAGACAATATGTAAACAAGAGCATGGCTGTGTTTCAATAAAACTTT
ATTTAAAAAAACAGTCAGGGACCGGATTTGGCCAAAGGCCATAGTGTGCC
AGCCCCAAGACTAGAGCAATGCACTTTTAACTTTTTTATTTTATTTTTGT
AAAATGCCAAGATCCACAAAAAATGCTATTGCACCCCGTGTGTTAGCACTG
TGACTCAAGGTTTGGGAAATTCTGCTTTGAAGGCGTGATAGACAGGAGAG
CATGGTCTGGCCCCTTGGTGCCTTTCTGGTTGCAGCGAGCATTTCAAACT
ACAGAGCAAGGCCAGTGGTCTGTTCAGCACTAGAGACATGCAGCAAGGTG
TCCTGGGGTGAGAAGATGCCATAACTGGTCCCCTTTCTATCTCCTTAGGT
CTTGGACTTCATTCCATTTTCTGTTGAGTAATAAACTCAACGTTGAAAAT
GTCCTTTGTGGGGGAGAACTCAGGAGTGAAAATGGGCTCTGAGGACTGGG
AAAAAGATGAACCCCAGTGCTGCTTAGAAGGTAAGGTTCTTGTAGAAATC
TACCTCAGGGCCAAAGTGTAATTCCTAGAGCAGAACTTTGCTAGGTGCTG
TGCACAGACCCAGTTGTTTCCTGCTGACTTGCACAGTAAGTGAGCTTTCA
AATTTCCCTGGACAAATAACTAGACAAGAGAAATTCTGGAAGAGAAAAGG
AAGCTTTGCTTCAGTGTCCAGGCACATCAGGTAGTAGATAAAAGGATCGT
CCTCACCTACAGATTTGGGGCTTTAGCATCCTGTTTGCCAACTGGATGGT
TGCATATGCTTCAAAATGCACCTCTTCCCTCCCAACATTCCCAAGTGGAA
GAGAAGCCTCCGATGAGAAGGAACTCTCTAAGGCTGGGCTGAACAAATGA
CCCAGGCACAGGGCATCTGAGTATTCCATGAGGAACACATTTGGGTGTTG
CCCATGGGGGACAATAGGAGGAGGCTTTTGACCCAAATGATTGTCTACTG
AGGTGTGACGGGAGAGGCCTGTGACATGCCAGAGGCCAAACCCGTGATCC
AGTTCATCTCTATTCTATGTTTCTGAAGAGGGAAGCTATGATTTAATGTC
ATTACTATCATGCTGCTCTAGTATTTCTCAGCACATACACAGAAGAGGGA
ATTAAATGGTCCTTGATACCCCTAAATCCTTGGAAAATCCGAATTGCATA
TGCTAACCTCACTGCGTCTGACTGCAGACCCGGCTGTAAGCCCCCTGGAA
CCAGGCCCAAGCCTCCCCGCCATGAATTTTGTTCACACAAGTAAGGCCTC
GGGGTGAGGTGATGGGGGTGGCTGAGGTGCGAGGGTGGGGATGGGGGATG
GAGCCATTGGGTCCTCTTACAGGGTGAGAGAATTGTAGAATGGGGACACC
TAAGGGTGCTGGATGGGGCTGAAGTCTTTCCTTTGTGGAAGCAAATCCCA
TTAGGAGATAACTCTGGGAAAGATGAGCCCGGGGAGGGGCAGGTGATGCT
CACCTGCTAAGAGGCAAAGGGCAAGGAAGAGTTTGTGCCTGGGAACCTTC
CAGGTGCCTCTTCTGACCATAGCCAAGAGACTGGAGACACAGACCTCCTC
CCAGCACTGAGGACAAACAGCCATGGGGCCAGTGGGGGTGCAGGGACACC
CACACCACTAAGGGCTCAGGGCGGCGCCTTCAGAGCCTGAACCTTCCTCT
CATGCTGCCATTTGAACACCACAACACCCTAATAGGAAACTGTTAACATT
GCCACTGTTCAGGTGTGGAAACCGAGACAGACAGTGGAGATTCCCTGCCC
TAGGTGACACAGGTAATAAGTGACAGATGTGGAAATTTAAAGGTACTATA
ACGTCTGTCTGCCTGACTCAGGCTTAAGGCTCCCATCACCTCCTCTTCTC
AGGACAGAGTCAGGAGGCCTCAGCCTGAGCCCCAGCTCTAGTGCAGGTTC

FIG. 4 (53 of 61)

ATGTGGGAATACTGAGCCTCACTAGTACAATGGCAGAGAGGACCAAATGG
GACCAGGTGTGTAAGGGTGCCTGGCACAGTTGGGGAGGCTGCTGTCGCT
TCTCCACCGCTGCTGCTGCAGTTACCTTTGATGTTTTAGTTTTGTTGTAG
TTACACCATTGCTGGCTTTGGATCTGCACTGTGTCCACTCCAGGTGGAAC
CACGCACACAAGCCTCTCTGTCGGCCTGTCCTGACTTCTCCTTGTCAGG
GCTGGGATCTCCTTCAAATCTGGCGGAAGTGGTTCTCCAAGTCTGGTCCT
CAAACGTCAGCAGCATCAGCGCCTAGAAGTGTTAGGAATACACATTCCCA
GGCCCCACCACAGACCTCCTGCCTCAGAAACTCAGGGCGCTGAGGCTCTA
GGGGCTGCTTTAACAAGCCTTCCAGGTTATCGTGACGCACCTTGAAAGTC
TGAGAGCTACTGCCCTACAGAAAGTTACTAGTGCCCTAAAGCTGGCGCTG
GCACTGATGTTACTGCTGCTGTTGGAGTACAACTTCCCTATAGAAAACAA
CTGCCAGCACCTTAAGACCACTCACACCTTCAGAGTGGCCTTGAGAAAGA
TTTGGGGTCAAGGATCATGAGCGAGAACACCACTTAAGAGGATAGTGAAC
TAGTCTGCATGTGAGACGCTGAGATCCTATGTCAGGCTGTGATAGGAGGG
AAACAGAAACCAAAGGAAAGAACAGCTTTAAGAAGCGCTTAAGAGGTACA
AAGTAAAATGATGGTGCTAGAAAAGTAGCTTCTTAAAAAGAGCATTTTCC
AGTCTCACCCTGGACTAACTGAATGAGAATCTCAGGAGTGTGAGGCCCAG
GTATCCATGGTCTTAAAATGCCACCCACCAGGTGATTCCCAGTGTGCACC
AGGGGTGAGAGTCACAGCCTTAGGCCATGCCACTCAAAGGGTGTCTTCAG
ACCAGCAGCACCCACAGCTCTGGGAGTGCATCAGAAAGACAGAGGCTTGG
CACCACCCACACCTACTGAACCATAGTTTGCAGGTGATTTCTTGCACATT
AAAGTGTGGGAAATGGAAAAGCTTAGAGTTCAGCTAGCTCGGTGACTCTC
AGTCAACCTGCACCTGCTCCATGAACTCAGACTGCCTGGGATGGGCCCAG
AAAAGCTCCTGAGGAGATTCTGATGTAAGGCAGGGCTGATAACCATGGAT
CTCATCTGACCCCATATCACTGGGGAGTTACTTAGGATCTTGCCTGGGGC
CAGTCATCTCTTCCATAGACACTGAGAGTGTCCACGATGCTTGGGGCACT
ACAGGGTGGGAGGTGGAGGATCACGGGTGAGTCAGATAGGAAGCCTGCTC
CTGGGGAGCTTACAGTGCTATAGGGCAGCAAGCCAAGGATGCCAATACCT
GTGTGCAGGTACCACTGACGAGTGCAGAGCGCTGCAGCACCAGAGAGGAA
GCTACCCTGTGCAGAGGGGGCTGAGGAGGGCTGCAGGGAGATGACAGGAA
AGCCGGTGTTACAGGAGGAGTCCTCCCCACTCTTTGGGCATGAGGAGACC
AGGAGGACATTCTACAGTGAGAAACCCAGGCAGAGGCCATGTGCTTATGG
CATGGGAAAAGAATGACACCTTAGACTTATTCTCTACATTAGAATTGCCT
ACCACAGATACCCATATTATAGCTTCACATAGTGTGGTGGTTACTGTGTT
TTCATATTGTCACATTTGCCATTTTCCAGCCACCCACCCATTCTTGACAG
TCACTGGCCCAGCCTGGGGGCCCCTGTTCTTTATCAAACAAGTGCCTGAG
CTCTTTGCAGAGGTGAGGGTCACCTGTCCAATCAGAGGCCAGGAGGGAAC
GTTCCCTTTTAAGACCCTACTCTAGGCAGGCCTGGCCCAAATGAGTTGCT
AGGAGCCCACGCCCTAAGAACCCTCTGAGCACTGTTGTGGCTGGTCCTGC
TGCTAGAAGTTGTTCCTCCAGGGCCAGGTGCAAGATTTGTGGCTTTTCAA
AGGAGCCACTAAAGCTCCAGCTCAGCCTTGCACGGTGCTGGGCTCCTGGG
GGCTTCCTGCCTCCAACCCTCCCAACTCTTCCATCACCGCTCCCTTAGCC
TGGCCAGTGCAGGGATCTGTTCCACTCTAGGCACTGCTGAGGGAATGATG
CCTCCAGTCAGAGGGTGCAAAAAAGAGAGTTAAGAAAAACAATGATTATA
AAAAGTCCTTTTTATACGCCAGACATTTTCTTTGCTCAGGCTAAGTGCTA
CTTATTTGAGTAAGCATTTTAGTTCTCATAACTCCTCTCTCAAGTAGGTG
CTGCTATTACTTTCATTTCACAGATGAGGACATTGAGGTTTGGAGAGACT
TAGTAACTTGTCCTCTGTCCTACAGCAGAGCTGGGATTTGAATCTATCTG
TCCAAATCTGGAACCCATTTGCTTGCACAGAAAGCTTAATTGCTTGTCCC
AGCAAGATAGAAAGCCTGGGAGTGGAAGAAATATTCAGTGGCTGTGATGT
CTGAGCCCACAGGCAGGGTGGAGAGCTAGGGCTGGGGCCCTTGGACGTGG
GGAAGAAAGGGCTGAGTCTTCCATTTTCAATGTGAAGTGTTGATATCTGG
TGATATTGATCTAGGTCCAAAGGTGAAGAACTTAAACCCGAAGAAATTCA
GCATTCATGACCAGGATCACAAAGTACTGGTCCTGGACTCTGGGAATCTC
ATAGCAGTTCCAGATAAAAACTACATACGCCCAGGTGACTCTCAGTTTTG
GCTGTGTTTTCTGCCTCCACCTAGCAGGGGTAAGGCCTCCTGCTAGGTGG
GCTCAACTCCATGCTATACCATGCCCCATCTCCAGCAGGTGGTGGAAGCG
AGGAGGAGAGGCCCCAGGGACTAGGGCATCAGATGAAGGGTCTCTAGCAA
TGACCAGATCTGAAAGTAGTCTTTCTGGAAGGGCTGGAGAAAAAGAAGGA
GGCAGACACTTAGACTGGAAGAAGAGGAGGCTTAAACCGGTGTGATGGAG

FIG. 4 (54 of 61)

GGAGAAGTGGACCACAGAGTCAAGGGAGAGGGACTGTGCATCAGGCCTGA
AACCCCAGCAGACAGGAGAGACCTTTCCCTGCTCTCAGAACCCACACATG
TTCTGACTGTCTTTTTCCAGAGATCTTCTTTGCATTAGCCTCATCCTTGA
GCTCAGCCTCTGCGGAGAAAGGAAGTCCGATTCTCCTGGGGTCTCTAAA
GGGGAGTTTTGTCTCTACTGTGACAAGGATAAAGGACAAAGTCATCCATC
CCTTCAGCTGAAGGTGAGAGTTCTAGCTCAGTTTCCTGGGCCTTTGGCTA
CCCCAAAGTAAAAGGCCAAGATCCTCAATGCCTCTCGCTTTCCTGCAAAT
TCTTATCTTGGCCAATATAACAGGGACATCCACCTTTCTGGAAGCACCAG
GCAGAAGAGCCCCATAACTTCTTCTCTGGTTCCTTGCCCCTTCTAGGGAA
GGAGGAGAGACTCCTCACAGCGGGGAGACAGCAAGGAGCTGAGCACCTGT
TCTCCTCTCCTGGGCTCACTGGTCCTGGCCCTGGGCGGGTGGCGGTCCCC
TCCTGCTGTGGCCCTCCATGTGGCAAGCAACACAATTGGGCCAGGACCCT
GGCGTGCTGCTGTAGGGTAGGAGGGTGTGAGGGAGCACTCGGAGGGCAGT
GTGTCTGCCCTGCAAATTTAGTCCTGGATGGAGCATCCTTTCACTTGAGG
GGAGAAATCTTAGGAAGCTGAATTAGATACAGATCTAAGCCATATTCTCT
AATTTTAAAAACTATAGAGCTGAGATTTTGGTATCCATCTGACTCTTACG
TCTCTCTCTCTCTCTCTCTCTCTCTCAGTTTATTTTTAATCTGGGGGACA
AGAAGGCCTGGAAAAGAGGGCATGATTGCTTATCATCCCTTAAATACCAG
TACCAAGGCTGACACGTCATCTTTCCCAAGGACCATCTGCCTTCTCTCTT
TTCCTCCTCTCCTGTGTAAAGGCCTGGAGGATGAGCACATGTGCTGTGTT
TTCCTCCCTCTCAAAGCCTGTGCTATCTAATTAATCCCTTTTACCTCACA
GAAGGAGAAACTGATGAAGCTGGCTGCCCAAAAGGAATCAGCACGCCGGC
CCTTCATCTTTTATAGGGCTCAGGTGGGCTCCTGGAACATGCTGGAGTCG
GCGGCTCACCCCGGATGGTTCATCTGCACCTCCTGCAATTGTAATGAGCC
TGTTGGGGTGACAGATAAATTTGAGAACAGGAAACACATTGAATTTTCAT
TTCAACCAGTTTGCAAAGCTGAAATGAGCCCCAGTGAGGTCAGCGATTAG
GAAACTGCCCCATTGAACGCCTTCCTCGCTAATTTGAACTAATTGTATAA
AAACACCAAACCTGCTCACTAAACTTTCTGTCATTGGGTTTCATTTCTCA
TTCATGCTTTAAGGATTTGTGTTTTTAGGATATAGCAAGAAGCTTGTTTA
ATTACAAAGTTCTGGGTTGGAAAGAGACCGGCTTCTGCTTGTGTACTGCT
ACCCTGAACCATCAGACATGCATGTGTGTGTCATATGCTATGATGTGGCC
AGTCTGAGTGCAATACTTGCAGCGGGAAGGAGCAGCTGGGTGCATGCTGT
GCTCTAGAATTAGTCTTTCCTACTGGGGTTTGGTAGATTCTGAGGGCATT
GATCCTGGGGCAGAAGTGGCTGAGTCTGTGTCTAGGGTACAGTGTGCAAG
AAAGAAATGTAACAGCAAGTCACAATCCAGCCAAGTGATAGTGGAAAAGG
GGTAGTTAGGTCCCAGATAAGGAGCAGGGTGACTTGACCTGTGGGAAAGG
CACAGAGACAAGGAATCTGGGTCAGATGACAGCCAGGAGACCAGGTGAGG
GAGGAGCCAGGTACTGTCTGGGAGGCTTGTCAACAAGGGCATGGTCCTAT
CACTAAGCAGGGCTCAGATCCTCATAATGGGGGAGTGGAAGGCTGGCCGA
ACAGAAATCAGGGCCTGGAAACAGAGTGAGGGGGTGGAGACAGGAGACTG
AGGCTTGGAAATTAGTTTATTAGTTTTAGCTCTTCAGTTACAAGCAATAA
TAATAGCTTCTAGCTTATTTAAGCAACAAGTATACTACAAAAGGAGCTTT
CTAGAAGGATATTGGGTATATTCATTTCTTACTGCTGCTGTAACAAATTA
CCACCAACTTAGTGGTTTAAACAATGCAATGTATTATCTTGCAGTTATGG
AGGTCAGTCTGGAATGTGTCTCACTGGGCCAAAATCAAAGTATCAGCAGG
ATAGCATTGCTTTGGGAGGCTCTAGGGGAGAGTCAATTTCCTTGCCTTTT
CCAGCTTCCAGAGGCCACCTGCATTCCTTGGCTAGTGGCCCACTCCCATC
TTCGCTGCTTGGGTTTTTCTCACACTGCTTTGCTCTGACCCTCCTGCCTT
CCTCTTTCACATATAAGAACGCTTGCAATTTACATCGGGCTCACGTCAAT
ATCCAGGATACTCTCCCGTCTCAAAGAGGCTTAACTTTAATCACAGATGC
AAAGTCCCTTTTGCTATGTCATGTAACATATACACAGGGTCTGGGGATTA
GAATGTGGACATTTTCGGGGTGCCATTATTCTGCCTATCATGTGAAGTAA
CTTTCAAAATGGAAAGACATGCTGAAGAAAAAGTCAGGGATTTCTGGCAG
GCCAGAAATGACAGAAGGCAGAAAACGTTGGTCCCATCACTCAGATGGGT
AAGAGCCAATCATGCTTTTTGTCAGTTAGCAAAAGATTGAGATTCCAAGC
AAAGCATGCAACTGCCCTAGTTTGGGTCATGTGTCGACTCCTTGGTCAGT
GAAGGGCAGCACACCTTGATCAATACTCCCTCCAAGACTGTATCCAACGA
GGCCAGTGATGTTCCTCAAAGCAGAGCTAGAGAGCTAATCCCAGGAGAGA
GGCGTGTGGGTGGTGGGCAGGAAGACAAAGCTCAGCCGTAAAGGAGTAGT
AGGGACAGCACCCTAGGCATGGAGGCTCAAGTGAGATGATACCCATGGGA

FIG. 4 (55 of 61)

AAAGCTCTGATAAGGTCAGCTCCTTCTGTTTCTGATCCTGATGGTGATGG
TGATCAACACCAGCCCAGTGACAAAAAAGTACATAGTATATTTAGTAGAT
GTTTCCCACACAGAGAAATGGTAAATATTCAAGGCGAGGAATACTCCAAA
CATCCTACCTTGATCATTACACATTCCGTGCATGTAATGAGTACTTGCAT
GTATGCCATAAATATGTGAAATATTATGTATCACTATATAAAAGAAAAAA
AAATGTGGCCAGGTGACATCCATATTTTGGAGAGGAAGGCATGTCTTCTT
CATAATATCACAAAACTATTTTCACAACAAAGACACAGCTGTTCAAATTA
GTCTCTGAGCCGGGGCTGTCTCATGGCAGTGAGGACTCTGGTTCCCTTAC
AGACTAGCAGAAAGGAGATGGGGCTTACTGACCATGGCCTTGAGGAGGCT
GAACATGCAGGCCAAATGGAGACACAGACAGCCTGGGCTTGGTCCTGCTC
CATCCCCTTCCAACCTGATGAGATATAGTGAGTCACTATGACGTGGGTCA
CTCATGCTTCCTGTGAGGCTCCACCAAGACAGCAAGTGCATCAACACCTT
ACGGAAGCACAAGGCCCTGTTTGTTGTTGACTTCATGAAAGGCATGGTTG
TGGTGATCGCATTGAGTAGGCTTTTGGGTGAGAGGTGAAAAACCCCAACT
ATCATGCATTGCAGCCCTCTGGTGGAAACTGTGCTTCAGGCTCTAAATTT
CAGGCTCTAGACTGACTCCAGGATGAGTATTTGGAAGCTGAAGTCAATCT
GTGGTCTCTTCTCCTGTAGAGCAGGAGTCAGCACTTTTCATAGAGTGCCA
GATTCTATATATCCTGCCACATGCTCTGTTGTTACAGAACAAAGAAGGCC
ATAGACAGCATGGCTGTGTTGGCAAATACACAAAACAGGCAATAAGCTGT
ATTTGGCCTTTAGGCTGCAGTTTGCCAACCCCTGCACTAACACAGAGCTT
AAAGGTGGTGGTGGTGTGCTGGAGCTAGCTTATATCAGCTTGCAATAGCC
AATTGCTAACATCTCTTCCAAACTCTGTGTCTGTGCCTTGATGTTGATAG
TTTGAAATTGGCTACCCCATTTAATGCTGCAATCTTTTCTCACCCCAGCA
CTACTGACTCCCCTTTGCCCTGTCTTATTTTTCTCACTCTAACATGCTGT
ATAGTTTTCTTCTTACATTTATTGTTTGTGTCTTCCACTAGCATGTATGT
CCCACAAGTTCTTTGCTCTGTGATGTATCCCAAGAACCCACTGCAGTGCT
TGGCACTTGTAGGAACTCCATAAGATTTTTATAAATGAAGAAAGGAAGAA
AAAAGAGAGGGAGGGAAAAAGGAAAGGAAGCCTTCTATTTAAATGATGGC
CTTCTCCATATTTCTATAGTAATATGACTTCCCTTGCAAAGGGGGATGCA
TTTTGGAAAATGTGTATAAATAAACTCAGGTGGTTTTGAATTTCATTTTC
CTAACTGTAATTGTAATCATTGGTCTTTATGTTTAGTGAAAAAGTTTTGG
CCCTTATGCCTCACACCTGAGAATCCCAAAGTATTGGTTTGTTAGAGCTC
CCATAGAGAACCATAAACTGGGTGGCTTAAAACAACAGAAATGTATCGTC
TCCTGGTTCAGGAGGCCAAAGTCTGAACTCCAGGTGTTGGTTCATTCTGA
GAGCTCTGAGAGAGAATCTGTTCCAGGCTTCCCTTCAGTTTGTGGTAGCT
CCAGGGTTCCTTGGCTGGTGGCAGCAAAACTCCAGTCTCTGCCCCCATCT
TCACATGACTGTCTTCTCTCTGTGTTTCTGTGTCCAGATTGTCCTATAAG
GACAGAGTCATACTGAATTAGGGCTCACTCGAATGACTTCATCTTAAGTT
GAACTGTATCTGTAAAGACCTTATTTCCAAGTAAGGTCACATTCACAGCT
ACTGGGGGATAGGACCTCAACATATCTTTTTGGGGGACATAATTCAACTC
ATAATACCCAACATGATAACTGTTCATCCCATGAAATTTAATGTCTCTCA
AAAGGTGATCTCAGGGCATTTAATCTGTGACAGAAACTCCCATAGGAAAC
ATTCCAACCAGAAGCTCCTTTCACAGCTGGTCACTCCTCCTACCCCATCC
GAGGTCCTGGGGCAGGGTGAGGCAGGTGGGGACAAGAAGAAGGCTGTCTC
GGGTGTAGAAAGAGAAGACCCTTATTCACCCGGCACTCTGTTCATGAATG
AGCTATCCAGCATAGGATATAATAAATCGCTTTAGGAGTGGTAGACTCCA
AACATTTTTTTGGTCCCAGTTATCCTAATCAATTAAACAAACTCTAGAAC
CCATCTTGAAGTGCAGGCATTGGGACATTATGAAACTTACACAGAATTCA
AAAATTTACAAGGGCTAAATAAAACAGGGTCTGACATCTAATATTTTCTT
CCCACATTCCCATGCACTGTCTGGCTCAACCATCCCCAACCCTCACTCTC
ATCCTGGTGGACACATGCCTAGTGATGTGATCAGCTGGTTCACAGGGGGC
TGGTGATGGTGGATATACAGCTTTTGCCAATTTCCATGGCATAACTACTC
CAAATATGGCCAATTTCAAACTACCAACATGAAGGCACAGACACAGAGTT
TGGAAGAGATGTTAGCAATTGGCTATTGCAAGCTGATATAAGCTAGCTCC
AGCACAGCACCACCGCTACCTTTAAGCTCCTTGTGTTAGTGCAAGGGTTG
GCAAACTGCAGCCTAAAGGCCAAATACAGCTTACTGCCTGTTTTGTGTAT
TTGCCAACACAGCCATGCTGTCTATGGCCTTCTTTGTTCTGTAACAACAG
AGCATGTGGCAGGATATATAGAATCTGGCAGTCTTTAATAAGTGCTGACT
CCTGCTCTACAGGAGAACACAGATTGTCTTCAGCTTCCAAACATTCATCT
CTGAGTCAGTCTAGAGCCTGAAATTTAGACTGAAGCACAGTTTCCACCAG

FIG. 4 (56 of 61)

AGGGCTGCAATGCATGAAAGTTGGGGTTTTCACCTCTCACCCAAAAGCCT
ACTCAATTTTTTACTGCAAAAACATGTTATCATCATTATTTTTTACTTAG
CCCACCTTTCCTTGGCAATTTTCCATAGGAAAATGCATTCTAAATTTCAA
CTAATCAGGGGACTTGGAGCCTCTGGACACCCCCTTGTTCCTTGCCCACA
GTCCCTTGCAGAAGGTGCCTTATCAGAGCGGCTCCATGCAGGGGCTCAGG
ACAGGATCAGATGTCAGTTGCACCAAGGGGGCAGGGACAGATCCTCTCTG
CTGACCATGCAGAAGGGACTGTTCAGTGCACCGTCATGGTCCTGGTGATT
TCTGGTCCATAAGGGAATTTTCACATGCATCGGGTGATTGTCACATCAGC
ACAACACTGTGAGGAAGGCAGAGTGAGAATTTGTGTGCCCATTTTATAGG
TGAGAAAACAGATGCAGAGACATTAAGTAACTTCACCACAGTCATGCGGG
TTTTAAGTGGCAGACTTTCAGGTGTTGTGACTCCTAGTCCAGAGTTCTTT
GCACTGCCCCTGAGGTGCTAAAACTCTACTGTGCTTTAAGACTCACTTGG
GGAGCTTCCTAAAAAGAGAGATTGCACAACCTGAGATTCTTGTTTAACTG
TTTTGGGATGTAGCTCAGGGATCTAGCTGCCTTAAAAAAAAAAACTCCCA
AGTAATTCTGATGCAAGCGGTTCTTTTTTGTCCACCTTTGAAGAAACACT
GCCTCCTCCCCATACATTTCATTAGAAAATGGTAACATGTTTTTCAGCCT
GAGAGCCATTTCTGGGTGACCGGACGTCGGCAGCCCGCTGTACTAGCTTT
CAGTCTAGGCTTAAACACACATGATAGGAGATGTCCTACTCCAGATGATA
TGAGTCTGAACCATGGAAAAATTCCATTGTGTGGCACATCTGGTGGGTGT
GCACTGTCCCCAGCAGTGAGGCACCCAGTGAAGACAGCAGCTGGGAGAGG
CTTAGTTACATGCAGTGGGACAGTGTGGGCTAGACTGCTGAGCCCCTCTGC
AGTTTACTCTGTGTCAGGCAATGAGGGTGAAAGGCTGATCAGACCCACGT
GCAGACCATACCCTCCAGGGAGACAGATATCAGTCAGGACAACCCCAAGT
GTAGCTGGAGAAGCAGTGCCCAGGTATGACCGGATGTGTATCCAACCAGG
AAATCTGCATATAAATATAAGAGGAGAAAATGAACAGATGTTGCTCTTAT
ATGTAGATATTTATGAAGAGCATATAATTTTGTTTTGTGTGTTTTAAGAA
GTTTATAAGTATGCCTTAAAAATGTATAGTATATACTGTAGGTATTTTTT
CCATTAGATATTTTGTTTTTCATACTTATCCACATTGACATTGTAGCAAC
AGTATAATATAACAACCTCCTCTACAAAAGCAGAAGGAAGTGAAGCTTTG
GAAGGAAGCACCCAGTGAGCTTGCCCCTTTCAGGTGGGTGCAGTGAGCAG
GAGTCAGTGAGGTTGAGATCCTTTGAGAGGAGGCAATCATTAACCAGGAA
ATCTGCACTGCATCCTGGCCACACCTAACCCTTGGACAATGGTGCTTGGA
GCGCCTTCCAGCTCTTAAGGCTTGCGATTTCTTTCTCTCACTCTTCACCC
ACGATGATTAAATCTTCTCCTACAGAGTTGGACAATAAAGCCTTGAGTTC
CTGCCTCCCCTGGTGTGATCACGAGGCATAGACATGGCCAGGAACATGTA
GGTGTCTTTGAAAGCTGAACAAGTTAGTAAATTTCAAACCTCATTTCACC
CACCAGTAAAATGGGAATAATAATAAACCTATTTTACATAGGGTTGACAA
GAGGAGTAAAGAGGGATTCAATGAAAGTTCGTTATTATCATTTGTAGTAG
CAGTGTTGATAATATCAACTGAAAGTTCATTATCATTATTAGTAGCAGTA
TTGATAACCCTCTTTTCTGTGCCTTCTCACTGGTGGGCCCAGGCCATCAG
CAATGCCCAGGGTGTCATGGATCTCTGCTGCATCGGGCACCAGCTGTGTC
AATGGTGAGAACAGTACAAGGGTGGGCAGGGCAAGGCAGGAAGCACCCAG
GAGCAGCAGCTTCATGGGGTGAAGATGTCAGGAGCTTAGGGACAGTCAGA
GCGGGTGTGCCTCCTCTTGTGGAGCCTTTCTGCGTGGGTAGGAACTGCTG
CAGCTGTGGCCATGGATTCACCTGAATATGGGTGGAATTAGGCATTCAGC
TGGGTTAGCTGTGCCTAGAAGGAGGAACTCTAAACTGAGAACTTGTCCCT
ATTGCCACCTCTGATAGGCAGATGATCCATCCATCAGTGGCTGAGCTGAG
GTGTGCATGGGGATGGGTAAGAGCCCACACACAGGGCTGATGACTGAGTC
TATTTAGAACAATAGATGTAAAATCTGATAATGTAAAATGTGATAGATTA
TTTTGTCAATTAGAAATGGTACCATATAATTATATATATACATAAACATG
TATACATATACACACATATACATGTGTGTATAAACACACACAGTATTGTC
CCCTACTCATTCCATAAACCTGATGCCTTTAGCTGGGATTCCCAGCTTTC
ACTCTCCTCTCTGTCATCTGCTGTCTATATCCTCCCCATCCTGTAATTCT
GGCTTATATGCCACTTCCTCCCTAAAGCCCTCCCTCAATCCCTTGCTGGA
AGTGACATTTTCCTCTTTGAGCTGCCCCTGCTTGTGCTTTGGTGAGGTCA
GCTGTATTGCAGTACCTTGTATTGTGGTTGTCACATCATCGTATAGAATT
AATTTCTGACACATTCCGTATTTTTCAAAGGGCCTAGTGTGGGGCTTTAA
CAGTAACTACGCCACCACGCCCAGTTAATTTTTTGTATTTTTGGTGGAGA
CAAGGTTTCACCATGTTGGCCGGGCTGGTTTCGAACTCCTGACTTCAGGT
GATCTGTCTGCCTCAGCCTCCTGGAGTGCTAGGATTGCAGGCATGAGCCA

FIG. 4 (57 of 61)

CTGCACCCAGCCACCTATCAAAATTTTAAGTGCCATTTTTATTTTTTATT
TTTTGTAGAAATGGACAAGCTGATCGCAAAATTCACATGGAATTGCAGGA
GGTTCCAAATAGCCAAAACAATCTTGAAAAAGAAGAACAAAGTTGGAGGA
TTTACACTTTCCAGTTTCAAGACTTAGCTCTTAGCTACAAAGCTACAGTA
ATCAGAACACTATGGTCCTGGCATAAGTGATGCTGGACAGGTGAGCCCCA
AAGTGGGACTTAACCTGTGAAGGTTCTTGGCCTTGCCCAGGAAGGAATTC
AAGGGCAAGCCAATGGGACAAGAAAACAGCTTTATTGAAGGGGCAGTATT
ACAGCTCCAGCCCTGTTACAGCTCCAGCCCTGTTACAACTCTGACTACTC
CTGCACAGAAGGGCTACCCTGTAGGCAGAGAGTAGCAACTCAGGGCAGTT
TTGCAGTCATTTATATCCACTTTTAACACATGCAGATTAAGGGACAATTT
ATGCAGAAATTTCTACGGAATTGGTAATAACTTTTGGGTCATGGAGTCAT
CATGGAAGGGGGGCGGGGAACTCCCTGGTGTTGCCATGATGACGGTAAAC
TGATATGGCGAACTGGTGGGTATGTCACATGAAAAGCTCCTTCCACCCCA
GCCCTGTTTCAATTAGTCCTCGGTTTGGTCCAGTGTCCAAGTCCTGCCTC
CAGAGTCAAGTCCCACCCCCTACCTCTTAAGGAGAGATGTAAATACATGG
AATAGAATTGAGAGTCCAGAAATAATCTCATACATCTATGATCAATTGAT
TTTCAGCAAAGGTGCCAAGACCATTCAATGAGGGAAAGAATCATATTTTT
TTCAACAAATGGTGCTGGATAACCACATGTGAAAGAATGCAACTGGGCCC
TTATCTCACACCATATACAGAAATTAACTCAAAATGGCTCAAACACTTAC
ATGTAAGAGCTAAAACTATAATATTCTTAGAAGAAAACAGGGATATATCT
TTATGACCTTGGATTTGCTGGCTGATTCTTAAATGACACTGAAAGCACAA
GCAACAAAAGAAAAAAAAATAGGTAAATTGGACCTCATCAAAATTTAAAA
CTTTTATGCTGGGTGCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGG
CAGGAGGATCTCTTGAGCCCAAGAAGCTGAGGCTACAGTGAGCCGAAATT
GTGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACCCTGTCTCGAATA
AATAAATAAACAAATATATAATTATAGATCTCTGGATCTTGCCTTCGGAG
ACTGACTCAACTAACTGGTCTGGGTGGAGCCCAGCCATTTGTATTTTTT
GAAAACTCTCCAAATGATTTTACTGTGCAGCCAAGGTTGAGAATCACTGT
ATCATAGGGTTGGACTCCTAACTGGAAACAGTTTGCACCATCAGGTGTCG
CAGCATTCTGATAATAGTTAAGCTTTCCTCCTAGATTTTCTGATATTAGA
TGAGTCATGTTTACAAGTTTTTACCAAGAGACAAACTATCTTTCTGCCCT
TACTTTCTCTCTTATACTATTCTAATCCCAGAACCCTTTGGAACTTCCAC
TGAGAGATGAATCTAGAAAGTGACTCTCTTGGCTACAACAGAGAGTAATG
TTGGCCTGTTTGTGCCAGATCCAGTTGGTGCTGGTGGTGGACAGCACCT
CCCTGAAATCCCCTCCTCTCCCGTCAGATTCAGTCCCCCATTTGCATCAC
GTACAATCATCACTATGGGTTTCTATTACCTTGCTAGGGCATTTGGAGGT
ACCATATATACCAACTATTAGTTTTGAGCCATGGTTCCCAAAGTGTGGAC
TGTAGGGCACCTCAGCACACTCACGAGGTGTCATGGGATATTTAAATATT
CTGAAGAAAACACAGTGACATCTGTCAGGCCCGTGAAAACCGTTGGCATT
AAATTGTCTCAACCCAATTGCTTAAGAAGCAGAACTGGCCAGGCACGGTG
GCTCACATCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCAC
GAGGTCAGGAGTTCGAGACCAGCCTGACCAACATAGTGAAACCCCGTCTC
TACTAAAAATATAAAAATTAGCCATGCATGGTGGCATGCACCTGTAACCC
CAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAAGCGG
AGGTTGTAGTGAGCCAAAATCGTGCCACTGCACTCCAGCTTGGGTGATAG
TGAGACTACATCTCAAAAAAAAAAAAAATGAGAGAGAGAGAGAGAAGCAGA
ACCATCAGGTGTTTCTTTTGGCTTAAAGTACTCTGTGAAGAAATTCCTGG
GACACGAAGGATACCATGAACTGAGAGATTTTGGGAACCTCTGCTTTAGA
AGCTGGAGGTAGCATTCCTTGGGCACAGTACTGCCTTGGGATCAGCAAAT
CCTTTTGATGGTGCATTTAGGTGTGGCAAGACAGCTCTTAGAGTGGGACC
GGGATGTGCTTGGAGACAGAGGGAACTAGATTGAGCTGCCCGATAAAGAC
ATGCCAGCCTGGCAGAGTGTAGTGACTCATGTCTGTAATCCTAGTGCTTT
GGGAGGCTGAAGTGGGAGGATTGCTTGAGGCCAGGGGTTTGAGATCAGCC
TGGGAAACAACAAGACCTCTACAAAAAAAAAAGAAAAAAAAAAATTAACCA
CATGTGGTGGCATGCACCTGTAGTCCCAGCTACCTGGCAGGCTGAGGTAG
GAGGATCACTTGAGCCCAGGAAGGTAAGGATACATTGAGCCATGACTGTG
CCACTGCACTCTAGCCTGGGTGACAGAAAGAGACTCTGTCTCAGAAATAA
ATTAAATAAATAAATAAATATATAGTGGCCATGACATCCCTAGAAAGACA
AGGTCCTGGGAATAGGTAGAAGCCAAGGGAAATGAGAAATGAGAGGGGGC
CCTGGAGCTGGAACTGGGGGAGCAGGATGGCCTCTGAGAAGTTCCTGATA

FIG. 4 (58 of 61)

GTGGTGTCACTGATGTGTCTGATGTTTAGTTGTAATTATTTGCTGGGCCC
CTGTCATCCCTCATATCTGATAGCTCTTTGCTAGTCAAAGTGTGGTCTGG
GGATCAGCGGCATCAGCATCACTTGAGAACTTGTTAGAGATGCAGAATCT
AGAGCCCCACCCGGGACCCAGAAACAGAGCCTGCATTTTAACAAGCTCCC
CAGGTGATTCTCACACACACTCGCATTTGAGAAGCACTGGGCTAGTTGAC
AGATTCTCAGGCATGGCTGACATTGAAATATCCAGGGAGCAGGCTTGGCA
TTAGGATGTTTAAAAGTCCTCCAGGTGTTTCTAAAGCCAGGTTTGAGGAA
TTACTGGGCTGATACAAATGTTTTGTGATGATGCTTTGTGTGTGTGTGTG
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAGGGAATTC
TGGGTCACTTGGCACCAACACAGGAAACAATGGAAATATGTGAGCCATGA
CAGAAAGGTCAGGAGATAAAAGAAATTAGTGACATGAGAGGTACTCCTCA
GGTGTTAGGAAAGAGGGTAGAGCAAACCAGGTTTTCCACCATATGTTGGA
TAGGGGGTCAAGTAAATTTCTACTTAAAAATTACAAACAGGGGCTGGGCG
CGGTGGCTCATGCCTGTAATCCCGCACTTTGGGAGGCTGAGGAGGGCGGA
TCACAAGGTCAAGAGATTGAGACCATCCTGGCCAACACGGTGAAACCGTG
TCTCCACTAAAAATACAAAAATTAGCTGGGCATGGTGGTGCGTGCCTTTA
TTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAG
GTGGAGGTTGCAGTGGGCCGAGATCGCACCACTGCAATCCAGAGCGAGAC
TGTGTCAAAAAAAAAAAAAAAAAAGAAAATTCCAAACAGGATGACCCTAAG
CCTGCAGGACTTGGAGACATCTAGGTGACTGATACTCAGTCACAAAACAT
AATTGGTCACAGGCCTGATGAAATGCACAGCAGACCTTCAGATGGTATGC
ACTCAAGTGATATCCACAAGTCCACCTAAAGAAATGCTATATTCAGACAT
TTGGCATCAATCTCTATCAAACAAAGATAGTCCAAAGCAATGGGTTCCAA
AAACACTTTCCTAAGACAAATTCTCTATTTGCTTTTAATATCAGTCATCC
CAGCCCTTGGAATAGAGGAGCAAATGATACCAGTGGTACCCTACCACAAT
GCACCAAGGTATTATACTCTCATGCTCCATTTTCTCCCTCTGTCTACATC
ACTAATAACTCATTGATTTCTGGTGCAAGCCCTGCTGGGAGAAAAAGTCT
ACTCTTGTACCTTGGAGCAAGTTGCTCAGAGTAGGTATCGAGGATAAAAT
TTGGAAAGTTAGAAAAGCTATTAGAAGGAGATCCTAGTAGTTGAAAACAC
AGCCTGGCCAAGTCAATGATGCTATTTCATCTCCCCAGCCTTGCATGTCC
ATAGCTAAGGAAGACAATTTAGGCTTGGGCTAGAGGATGGAAAGGGCAA
AATTACTGATGCCACAGCCCAGAGAGGTATTCTAGTAATCTGAGGGTGAG
GACCACATACCTGGTTCAGGGACGTACAGTGTTGACAGCTGTGAGTGGAT
GCCTGGAGTTCTGGCGTGTCTTCTAGCACAATGATACCTGAGACTCTTGC
ATCATTGGGAATAATAAAATGGGAGTGGATAGATATGAAATTATGATGGC
AATAAGCAATCAGCTAATAGCTTCATTGATGGGACAGATTAAAGATGGCT
GCAAATCCTTTGGTCCAGGTTTGGGATATAGGCAGCATTTGTATTGGAAT
GCTGATAGTCTGAGGCCATGAAAAGTCCACCTGCAGTAGTGGTAGGAGGA
ACAAGCCTCACTTTCTTCAATGTGTGTGACTGCTGTCTTGATTCCCTGGG
TGGCCAGTTCCATTCGTGTGGTTCTTTGGTCCACTTGACTCTGGGGTGGC
TCTGTGATGGCTTGACCAATACAATGTAGTGGAAATGATGCTGTCATCAT
TTCCAGCCTCTTCCAGCCTTAAGGAACTGGCAACTTTTATTTCTGTCCCT
TGGAATACTTGTTCTTGCAACCCATCCATCATACAGTGAGAAATTCTAAG
CTGCCCCATTAAGAGGCCCACATGGTGATAAATTGGGGTCTTACATACAG
CCCTAGCTGTGCTCCTAGCTGACAAACAGTAGCAACTTGTCACCAGGCGA
GTGAACCACTTAGGACTGTATACTCCAGCCCCAGTTGAGCAATGTGGAAC
AGAGTAAACCATCTCAGCTTAGCCCTGCCCAAACTGCAGAATTATGAGCA
AAATAATCCCCTAGGCTTTGGGCTGATTTGTTCCAGATTACTGGAACAGA
ATTTGGTACCAGGGGTGAGGTGCTACAGCAATGAAAGCTTAAGACACGTG
ACTTTGGTTTTGGGTCTGAGTGGCAGGGGAACTTGGCAGGCCTCAAGGAA
ACTTTTAGGGAGGGTTGAAGCATAGTGAGGAAAACAGTAGGGGAAGCTAG
AGGAAAAAAATGATGCTTGGTATGTAGTGGTGGGAAGTTTAGCAAAACTCG
CCTGATGTAATGTGGGAAATTGTAAGAACTCAGAACGATTTAAGGGCATG
TTTTATAGGTCCTTTAAGAAACTTCTAGGCCAGGCGCAGTGGCTCATGTC
TGTAATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACAAGGTCAGG
AGATCGAGACAATCCTGGCTAACATTGTGAAACCCCGTCTCTACTAAAAC
TACAAAAAAAAAATTAGCCGGGCATGGTGGCGGGTGCCTGTAGTCCCAGCT
ACTAGGGAGGCTGAGGCAGAAGAATGGCGTGAACCTGGGATGTGGATCTT
GAAGTGAGCCCAGATTGTGCCACTGCACTCCAGCCTGGCAACAGAGTGA
GACTCCGTCTCAAACCGAAAAAAAAAAAAAAAAAAAAAGAAACTTCTAGGGC

FIG. 4 (59 of 61)

TGGTCCCGTGGAAGCCTCACACATGGTACACAAAGGCTGTCTTGAAAAGA
AACGTAAGTGTGTTTTTTGGTTTAATAAAATTGATTATAAATGGATAATG
CAAAACATTTTAAAGAATTTTACTAGCTTACATTAGCAGATTTGGATCCA
GTGATTGTTACATTCTGGTACTGAGCCCCTGAATTACTTCTTTGAGTAAG
GCATTATACCAAAGCTATTGATAGTTGGGCTTATAGGGTGTATGTTTGAA
GAACTACTAATGTCAAAACCAATATTTCACGGTCGACAAGAGGACATCAG
AACTGGTAATCCTTATTACCATGACTGGCTGGACAGAATACTCAATGTAA
TGGGATTTCCTGCAAATAAAGACGGGGAAGATGTAAAAAAGATGCCTGAA
CATTCAACATTAATGAAAGATTTCAGAAGAAATATGTATACTAACTGCAG
CCTTATCAAGTATATGGAAAAACACAAAGTTAAACCAGATAGTAAAGCAT
TCCACTTGCTTCAGAAGTTTCTTACTATGGACCCAATAAAGTGAATTACC
TGAGAACGGGGTCCCTGTTTCTTCGAAGACCCACTTCCTACATCAGACGT
TTTCAACAGTTGTCAAATCCCCTACCCAAAATGAGAATTTTTAACAGAAG
AAGAACCTGATGACAAAGGAGCCAAAAAGAACCACCACCGGCAGCAGGGC
CATAACCACACGAATGGAACTGGCCACCCAGGAATCAAGACAACGGTCAC
ACACAGGGACCCCCGTTGAAGAAAGTGAGGCTTGTTCCTCCTACCACTAC
CTCAGGTGGACTTTTCACGGCCTCAGACTATCCGCGTTCCAATCCACATG
CTGCCTATATCCCAACCCTGGACCAAGCACATCCCAGCCGAAGAGCAGTG
TAGGATACTCAGCTACCTCCCAGCAGGCTCCACAGGACCCACGTCAGACA
CACGGGTACTGAGCTGCATCGGAATCTTGTCCGTGCACTGTTGTGAATGC
TGCAGGGCTGACTGTGCAGCTCTCCGTGGGAACCTGGTATGGGCCATGAG
AATGTACTGTACAACCACACCTGCCCAGTAGCCAAGTTCCTTCCACCGCT
TTTCACAGATCGGGGTAGTGGCTTCCAGTTTGTACCTATTTTGGAGTTAG
ACCTGAAAAGAAAGCGCTAGCACAGTTTGTGTTGTGGATTTGCTACTTTC
ATAGTTAACTTGACCTGGCTCAGACTGACCAGTACTTTTTTTCCGTGAC
AGTCTATAGCAGTTGAAGCTGAGAATGTGCTAGGGGCAAGCGTTTGTCTT
CATATGTCATGAATTCCTCCAGTGTAACAACATTATCTGACCAATAGTAC
ACACACAGACACAAGGTTTAACTGGTACTTGAAAACATACAGTAGGTGTT
AACTCAGTGAAATAACCAGGACTCAAAGTAAGATTATTTTGGTACACCTT
TCTTGTTAGTGTCTTATCAGTGAGTTGATTCATTTTCTACATTAATCAGT
GTTTTCTGACCAAGAATATTGCTTGGATTTTTCTGAAAGTACAAAAAGCC
ACATAGTTTTTTTCAGAAAGGTTTCAAAACTCCTAAAGATTAATTTCCAA
GTATAAGTTTGTTTTTATTTTCAATCTATGACTTGACTGGTATTAAAGCT
GCTATTTGATAGTAATTAGATATATTCTCATTGATATAAACCTGTTTGGT
TCAGCAAACAAACTAAAATGATTGTCACAGACAATGCTTTATTTTTCCTG
TTGGTGTTGCTTGTGGAAAAAGAAAGAGAGATCAGATTGTTACTGTGTC
TGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAAGA
AAAATTCTTCTGCCTTGAGATGCTGTTAATCTATATAACCTTACCCCCAA
CCCTGTGCTCTCTGAAACATGTGCTGTGTCCACTCAGGGTTAAATGGATT
AAGGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATG
CTCGTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAA
CACTGCTGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCC
AAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAGGGGAA
AGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGA
GGATTAGTATACGAGGAAGGAACGCCTCTTTGCAGTTGAGACAAGAGGAA
GGCATCTGTCTTCTGCCCGTCCCTGGGCAATGGAATGTCTCGGTATAAAA
CCCGATTTTATGTTCCATCTACTGAGATAGGGGAAAACCACCTTAGGGCT
GGAGGTGGGACATGCGGCAGCAATACTGCTCTTTAAGACATTGAGATGTT
TATGTGTATGCATATCTAAAGCACAGCACTTAATTCTTTACCTTGTCTAT
GTTGCAGAGACCTTTGTTCACGTGTTTATCTGCTGACCTTCTCTCCACTA
TTATCCTATGACCCTGCCACATCCCCCTCTCCGAGAAACACCCAAGAATG
ATCAATAAATACTAAGGGAACTCAGAGGCCGGCGGGATCCTCCATATACT
GAACGCTTGTCCCCTGGGCCCCCTTATTTCTTTCTCTATACTTGGTCTCT
GTGTCTTTTTCTTTTCCAAGTCTCTCGTTCCACCTAATGAGAAACACCCA
CAGGTGTAAAGGGGCAACCCACCCCTTCATTGCTGATTTGTGAGCGTGCT
TTAAGGTGAAAAAAGCATGAATGTTAACTTCCTTAAAAAGGTACAGCATC
CAATTCAAATATTTTTGTCCTGATTTTAATGCTAGTTGATGTAGTGCTAT
TAAAATTTTGTTCAACATGGACACAGAGAGGGGAACAACACATACCAGGG
CCTGTTGCGGGGTGGGGATGAGGGGAGGGAACTTAGAGGACAGGTGAACA
GGTGCAGCAGATCACCATGGCCCACATATACCTATTTAACAAACCTGCAC

FIG. 4 (60 of 61)

GTTCTGCACACGTATCCCATTTCTTTTTTTTTTTTAAGAAATAGAAAAAAA
AATAAAATTTTGTTCACTGATTCTTCCATTTTAAAACTTGTTTGCATGTG
GTTTAGGATGCCCTTACTTCAGCAAAGGAGAAGGAATAGGAGGGCCTTAG
AATTTTTGAGGGAAAAAAACCCTATAACATACATTGTACTGTATCAAACT
ATTTTACATGAATGACACAAGTATTCTGAATAAAAAAATAATTGAACATT
GTTAAGAACAAGGTGTCATGTAATTTATTTTTCATAAATAAAAAAATTAT
AGTGGCTTAGACTGAAAGGAACAGAGAATTTAAAAAATTAAAAAGAAGCC
TTAGTATATTTTTGTATATAGTTTCCATGTGCCATATTTGCCATAATTGG
ATGAGAATTTTTTGACCTCTGGCAGGGTGACCCTATATTTTCANTNTATA
AAGCGTGCATCATACC

FIG. 4 (61 of 61)

MVLKCHPPGDSQCAPGVRVTALGHATQRVSSLQQHPQLWECIRKTEAWHHPHLLNHSLQPGGPCSLSNKCLSSLQRSASA
EKGSPILLGVSKGEFCLYCDKDKGQSHPSLQLKEKLMKLAAQKESARRPFIFYRAQVGSWNMLESAAHPGWFICTSCNCN
EPVGIXNXVDFDLLGKAQKRGTGSE

FIG. 5

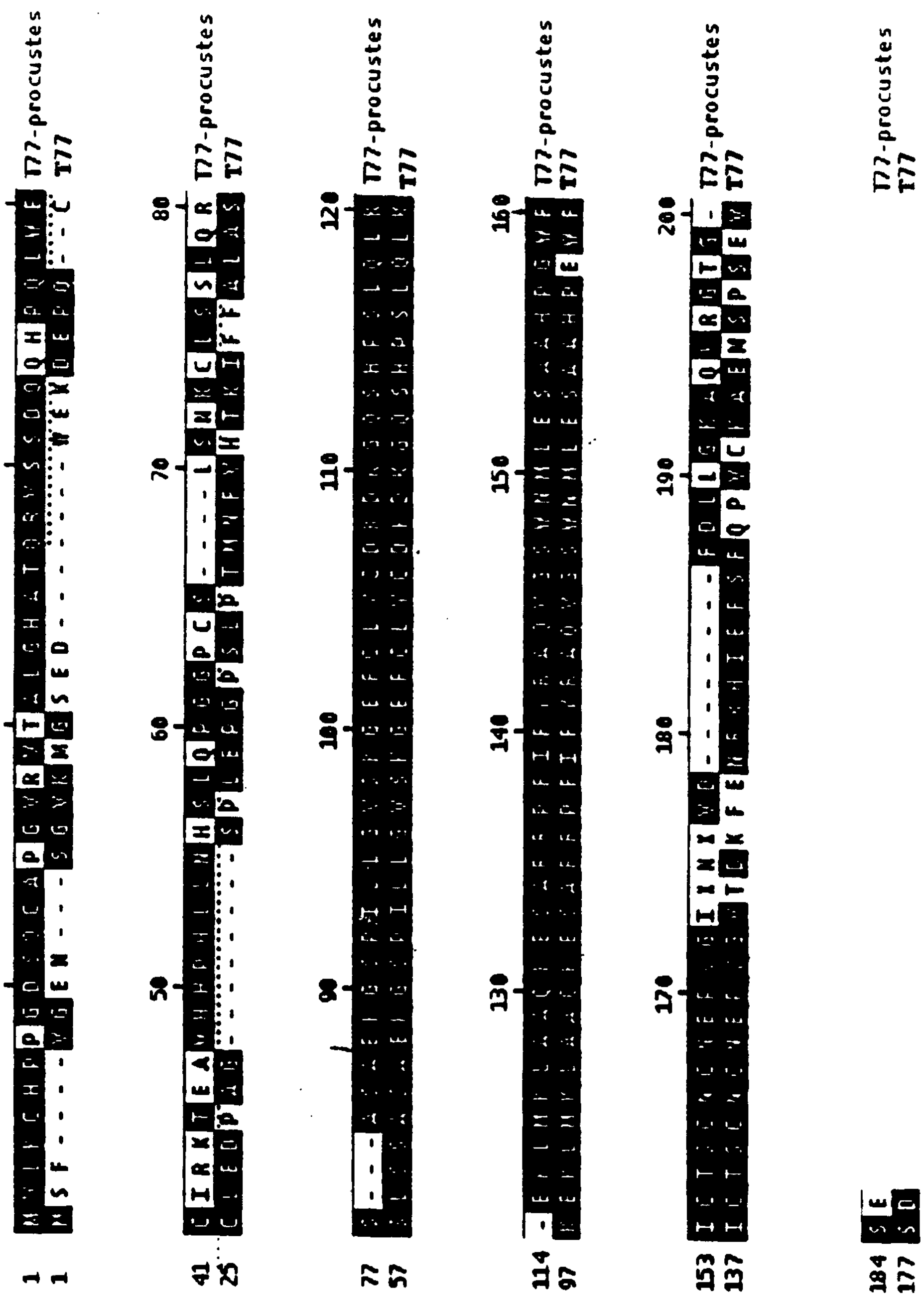

FIG. 6

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/16102

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07H 21/02, 21/04, 1/00, 14/00, 17/00; C12Q 1/68; G01N 33/53
US CL : 536/23.1; 530/350, 387.1; 435/6, 7.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1; 530/350, 387.1; 435/6, 7.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG: MEDLINE, USPATFUL, WPI, BIOSIS. Search terms include author, "TANGO" and protein

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Database Medline on Dialog, US National Library of Medicine, (Bethesda, MD, USA) AN 09370320. SONNENFELD et al. 'The Drosophila tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development'. Development. November 1997, volume 124, number 22, pages 4571-82, Abstract.	1-22

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 OCTOBER 1998	30 OCT 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
HEATHER BAKALYAR
Telephone No. (703) 308-0196